PCT WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 7 : C12N

A2

(11) International Publication Number: **WO 00/24870**

(43) International Publication Date: 4 May 2000 (04.05.00)

(21) International Application Number: PCT/US99/24973

(22) International Filing Date: 22 October 1999 (22.10.99)

(30) Priority Data:
60/105,324 23 October 1998 (23.10.98) US

(71) Applicant *(for all designated States except US)*: UNIVERSITY OF VIRGINIA PATENT FOUNDATION [US/US]; Suite 1–110, Towers Office Building, 1224 West Main Street, Charlottesville, VA 22903 (US).

(72) Inventors; and

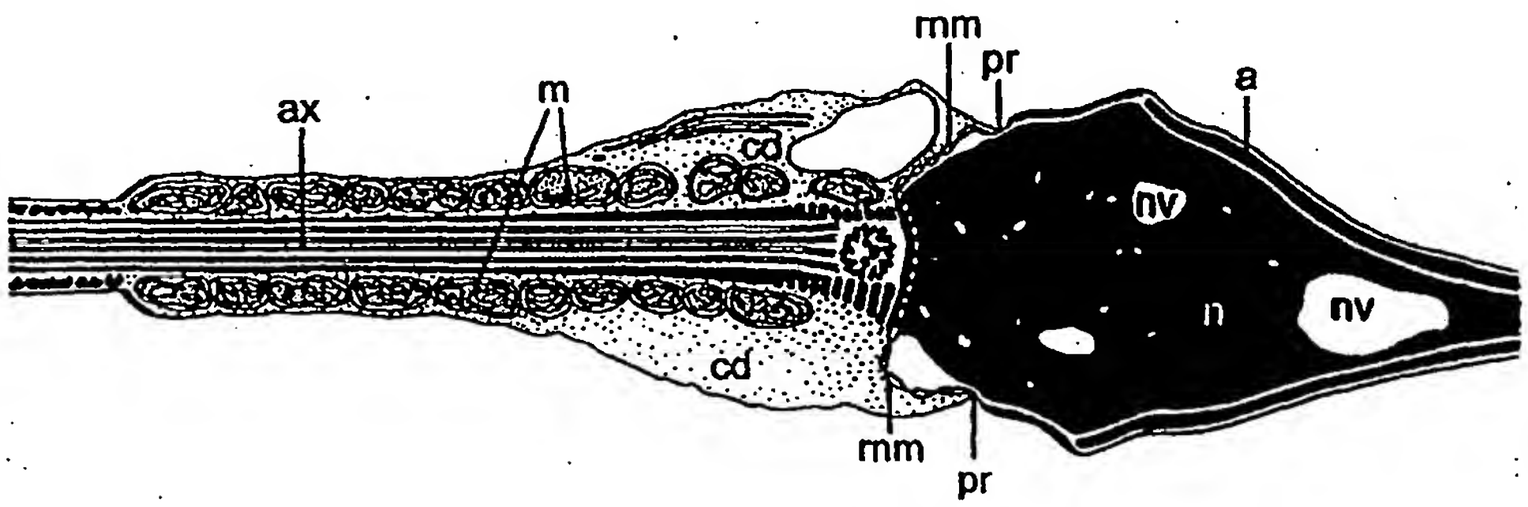
(75) Inventors/Applicants *(for US only)*: HERR, John, C. [US/US]; 2545 Cedar Ridge Lane, Charlottesville, VA 22901 (US). WESTBROOK, V., Anne [US/US]; 500 Druid Avenue, Charlottesville, VA 22902 (US). DIEKMAN, Alan, B. [US/US]; 2309 Highland Avenue, Charlottesville, VA 22903 (US).

(74) Agents: BALDWIN, Geraldine, F. et al.; Pennie & Edmonds LLP, 1155 Avenue of the Americas, New York, NY 10036 (US).

(81) Designated States: AE, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CR, CU, CZ, DE, DK, DM, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, UA, UG, US, UZ, VN, YU, ZA, ZW, ARIPO patent (GH, GM, KE, LS, MW, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

Published
Without international search report and to be republished upon receipt of that report.

(54) Title: NUCLEIC ACIDS AND PROTEINS OF THE HUMAN SPAN–X GENE AND USES THEREOF



(57) Abstract

The present invention provides the discovery, identification and characterization of nucleotides that encode a novel X–linked spermatozoan nuclear protein, SPAN–X. The invention encompasses nucleotides encoding SPAN–X, proteins encoded by the nucleotides of the present invention, and antibodies and ligands that bind SPAN–X as well as fragments of any of the above having a functional activity of SPAN–X, *e.g.*, a nuclear localization signal activity, etc. The present invention provides screening assays to identify SPAN–X–bearing spermatids. The invention further provides compositions and methods for the use of *span–x* nucleic acids for heterologous gene expression.

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	**ES**	Spain	**LS**	Lesotho	**SI**	Slovenia
AM	Armenia	**FI**	Finland	**LT**	Lithuania	**SK**	Slovakia
AT	Austria	**FR**	France	**LU**	Luxembourg	**SN**	Senegal
AU	Australia	**GA**	Gabon	**LV**	Latvia	**SZ**	Swaziland
AZ	Azerbaijan	**GB**	United Kingdom	**MC**	Monaco	**TD**	Chad
BA	Bosnia and Herzegovina	**GE**	Georgia	**MD**	Republic of Moldova	**TG**	Togo
BB	Barbados	**GH**	Ghana	**MG**	Madagascar	**TJ**	Tajikistan
BE	Belgium	**GN**	Guinea	**MK**	The former Yugoslav Republic of Macedonia	**TM**	Turkmenistan
BF	Burkina Faso	**GR**	Greece			**TR**	Turkey
BG	Bulgaria	**HU**	Hungary	**ML**	Mali	**TT**	Trinidad and Tobago
BJ	Benin	**IE**	Ireland	**MN**	Mongolia	**UA**	Ukraine
BR	Brazil	**IL**	Israel	**MR**	Mauritania	**UG**	Uganda
BY	Belarus	**IS**	Iceland	**MW**	Malawi	**US**	United States of America
CA	Canada	**IT**	Italy	**MX**	Mexico	**UZ**	Uzbekistan
CF	Central African Republic	**JP**	Japan	**NE**	Niger	**VN**	Viet Nam
CG	Congo	**KE**	Kenya	**NL**	Netherlands	**YU**	Yugoslavia
CH	Switzerland	**KG**	Kyrgyzstan	**NO**	Norway	**ZW**	Zimbabwe
CI	Côte d'Ivoire	**KP**	Democratic People's Republic of Korea	**NZ**	New Zealand		
CM	Cameroon			**PL**	Poland		
CN	China	**KR**	Republic of Korea	**PT**	Portugal		
CU	Cuba	**KZ**	Kazakstan	**RO**	Romania		
CZ	Czech Republic	**LC**	Saint Lucia	**RU**	Russian Federation		
DE	Germany	**LI**	Liechtenstein	**SD**	Sudan		
DK	Denmark	**LK**	Sri Lanka	**SE**	Sweden		
EE	Estonia	**LR**	Liberia	**SG**	Singapore		

NUCLEIC ACIDS AND PROTEINS OF THE HUMAN SPAN-X GENE AND USES THEREOF

This application claims priority under 35 U.S.C. §119 (e) to United States Provisional Application Serial No. 60/105,324 filed October 23, 1998, the entire contents of which are incorporated herein by reference in its entirety.

This invention was made with government support under grant numbers HD 29099, P30 28934, T32 DK 07642, T32 HD 07382, U54 HD 28934, HG00333 (T.B.S.), awarded by the National Institutes of Health, and D43 TW/HD 00654 from the Fogarty International Center, and a grant from the Andrew W. Mellon Foundation. The government has certain rights in the invention.

1. FIELD OF THE INVENTION

The present invention relates to compositions comprising "*span-x*" (sperm protein associated with the nucleus on the X chromosome) polynucleotides, polypeptides, and derivatives and analogs thereof. Recombinant SPAN-X proteins, fragments comprising functional domains, derivatives, antibodies, and ligands are also provided. The invention further provides methods and kits and for the use of said compositions in screening, diagnosis, and therapy.

2. BACKGROUND OF THE INVENTION

The ultimate function of the spermatozoon is to deliver the paternal chromosomes to the egg. To accomplish this, the spermatozoon must undergo a series of processes including conservation of the paternal haploid genome, transport to the site of fertilization, recognition of the egg, penetration through the egg investments, fusion with the oolemma, and activation of the egg (reviewed by Yanagimachi, R. (1994). Mammalian fertilization. In "The Physiology of Reproduction" (E. Knobil, and J. D. Neill, Eds.), 2nd ed., pp. 189-317. Raven Press, Ltd., New York). The spermatozoon has a highly specialized, streamlined morphology with various structural components tailored to specific functions (reviewed by Eddy, E. M., and O'Brien, D. A. (1994). The spermatozoon. In "The Physiology of Reproduction" (E. Knobil, and J. D. Neill, Eds.), 2nd ed., pp. 29-77. Raven Press, Ltd., New York). For example, the sperm nucleus is organized and shaped for efficient transport and delivery of the paternal DNA to the oocyte. The nuclear chromatin mass consists of DNA coupled with heavily cross-linked, basic nucleoproteins, the protamines, that effectively neutralize the DNA charge and allow for its tight compaction

(reviewed by Ward, W. S. and Coffey, D. S. (1991). DNA packaging and organization in mammalian spermatozoa: comparison with somatic cells. *Biol. Reprod.* 44, 569-574). DNA packaging and organization in mammalian spermatozoa: comparison with somatic cells. *Biol. Reprod.* 44, 569-574. DNA packaging and organization in mammalian spermatozoa: comparison with somatic cells. *Biol. Reprod.* 44, 569-574. ; Dadoune, 1995). The compact structure of the sperm nucleus is thought to protect the male genome from physical, chemical and mutagenic injury during epididymal storage and transport through the male and female reproductive tracts.

Sperm nuclear shape is highly species-specific and, within a given species, typically shows a very high degree of uniformity. However, human spermatozoa exhibit variations in nuclear morphology that may be linked to heterogeneity in chromatin condensation, areas of poorly condensed chromatin, and the presence of numerous nuclear vacuoles (Pedersen, H. (1969). Ultrastructure of the ejaculated human sperm. *Z. Zellforsch.* 94, 542-554); Zamboni, L., Zemjanis, R., and Stefanini, M. (1970). The fine structure of monkey and human spermatozoa. *Anat. Rec.* 169, 129-154). Nuclear vacuoles are non-membrane bound cavities of irregular outline and size randomly distributed throughout the condensed chromatin (Zamboni, L., Zemjanis, R., and Stefanini, M. (1970). The fine structure of monkey and human spermatozoa. *Anat. Rec.* 169, 129-154; Zamboni, L. (1991). Physiology and pathophysiology of the human spermatozoon: the role of electron microscopy. *J. Electron Micros. Tech.* 17, 412-436). Physiology and pathophysiology of the human spermatozoon: the role of electron microscopy. *J. Electron Micros. Tech.* 17, 412-436.; FIG. 1, nv). Nuclear vacuoles have been identified in sperm of various mammalian species but are particularly numerous and large in human spermatozoa. Various nuclear and cytoplasmic elements may be sequestered within the nuclear vacuoles including irregular membranous structures and fine granular material (Pedersen, H. (1969). Ultrastructure of the ejaculated human sperm. *Z. Zellforsch.* 94, 542-554; Fawcett, D. W. (1970). A comparative view of sperm ultrastructure. *Biol. Reprod., Suppl.* 2, 90-127; Zamboni, L, Zemjanis, R., and Stefanini, M. (1970). The fine structure of monkey and human spermatozoa. *Anat. Rec.* 169, 129-154). In addition, fibrils and dense fibrillar structures located within the nuclear vacuoles are believed to be remnants of the spermatid nucleolus and are often associated with the condensed chromatin (Czaker, R. (1985). Ultrastructural observations on nucleolar changes during mouse spermatogenesis. *Andrologia* 17, 42-53; Czaker, R. (1987). Relative position of constitutive heterochromatin and of nucleolar structures during mouse spermatogenesis. *Anat. Embryol.* 175, 467-475; Sousa, M., and Carvalheiro, J. (1994). A cytochemical study of the nucleolus and nucleolus-related structures during human spermatogenesis. *Anat. Embryo.* 190, 479-487).

Within nuclear vacuoles, cytochemical staining has indicated the presence of ribonucleoproteins (RNPs) and deoxyribonucleoproteins (DNPs) that have been hypothesized to facilitate the rapid re-initiation of paternal chromatin transcriptional activity after sperm incorporation into the oocyte (Dadoune, J. P. and Alfonsi, M. F. 1986. Ultrastructural and cytochemical changes of the head components of human spermatids and spermatozoa. *Gamete Res.* 14, 33-46; Sousa, M., and Carvalheiro, J. (1994). A cytochemical study of the nucleolus and nucleolus-related structures during human spermatogenesis. *Anat. Embryo.* 190, 479-487; Biggiogera, M., Martin, T. E., Gordon, J., Amalric, F., Fakan, S. (1994). Physiologically inactive nucleoli contain nucleoplasmic ribonucleoproteins: immunoelectron microscopy of mouse spermatids and early embryos. *Exp. Cell Res.* 213, 55-63). Although not unique to nuclear vacuoles, ATPase activity was demonstrated in nuclear vacuoles by cytochemical methods Sosa, A., L. Calzada, S. Alva, and A. Gonzalez-Angulo. (1979). Distribution of ATPase in isolated human spermatozoa nuclei: a high resolution cytochemical study. *Int. J. Fertil.* 24, 125-129).

In addition to heterogeneity within the sperm nucleus, ultrastructural differences in the nuclear membrane have been observed proximal and distal to the posterior ring. Anterior to the posterior ring, the inner and outer lamellae of the nuclear membrane are closely apposed, largely obliterating the lumen of the perinuclear cisternae. In this region, the nuclear membrane is tightly associated with the nuclear chromatin and lacks nuclear pores (Fawcett, D. W. (1970). A comparative view of sperm ultrastructure. *Biol. Reprod., Suppl.* 2, 90-127; Curry, M. R., and Watson, P. F. (1995). Sperm structure and function. In "Gametes: The Spermatozoon" (J. G. Grudzinskas and J. L. Yovich, Eds.), pp. 45-69. Cambridge University Press, Cambridge). Distal to the posterior ring and embedded within the cytoplasmic droplet, the nuclear membrane evaginates forming the redundant nuclear membrane and enclosing a nuclear space devoid of condensed chromatin (FIG. 1, rnm). The redundant nuclear membranes have a wider apposition between their lamellae and have numerous, regularly-arranged nuclear pore complexes (Fawcett, D. W. (1970). A comparative view of sperm ultrastructure. *Biol. Reprod., Suppl.* 2, 90-127; Pedersen, H. (1972). The postacrosomal region of the spermatozoa of man and *Macaca arctoides. J Ultrastruct. Res.* 40, 366-377; Curry, M. R., and Watson, P. F. (1995). Sperm structure and function. In "Gametes: The Spermatozoon" (J. G. Grudzinskas and J. L. Yovich, Eds.), pp. 45-69. Cambridge University Press, Cambridge). The folds of redundant nuclear membrane form during spermatid nuclear elongation as a result of the reduction in nuclear volume that occurs due to chromatin condensation (Franklin, L. E. (1968). Formation of the redundant nuclear envelope in monkey spermatids. *Anat. Rec.* 161, 149-161). The functional significance of the redundant nuclear membrane remains unknown, although the increased

surface area of the redundant nuclear membrane may allow for increased nuclear-cytoplasmic exchange (Dadoune, J. P. (1995). The nuclear status of human sperm cells. *Micron.* 2, 323-345) and the abundant phospholipids may serve as an energy source for the mature spermatozoon (Mann, T. (1967) Sperm metabolism. In "Fertilization" (C. B. Metz, and A. Monroy, Eds.), Vol. 1, pp. 99-116. Academic Press, New York). Although these hypotheses are tantalizing, the functional roles of the redundant nuclear membrane and the nuclear vacuoles in mature spermatozoa remain unknown largely due to the lack of unique biochemical markers for these structures.

Nuclear localization signals (NLS) are generally short stretches of 8-10 amino acids characterized by basic residues as well as proline. NLS sequences are retained in the mature protein, may be found at any position as long as it is exposed on the protein surface, and can be present in multiple copies. Proteins enter the nucleus through nuclear pores by a two-step process: the first step is a rapid, signal-dependent binding to the nuclear pore periphery, while the second step is a slower, ATP-and temperature-dependent translocation across the pore (Garcia-Bustos, J., et al.,1991, Biochim. Biophys. Acta 1071: 83-10.; Silver, P. A. (1991) Cell 64: 489-497).

Precedents for the incorporation of nuclear targeting signals within a DNA-binding domain include fos (Trainer, I., and I. M. Verma 1991 Oncogene 6: 2049-2053);the progesterone receptor, in which the second finger but not the first functions as an NLS (Guiochon-Mantel, A. et al., 1991, EMBO J. 10:3851-3859.); GAL4 (Silver, P. A. et al., 1984, Proc. Natl. Acad. Sci. USA 81: 5951-5955); and the homeodomain proteins .alpha.2 and Pit-1/GHF-1 (Hall, M. N. et al., 1990, Proc. Natl. Acad. Sci. USA 87: 6954-6958; Theill, L. E. et al., 1989, Nature 342: 945-948). If nuclear localization signals and Cys.sub.2 His.sub.2 finger domains--both typified by basic residues--have co-evolved, NLS sequences may generally be found adjacent to or integrated within zinc finger domains.

Other bipartite nuclear localization signals have been characterized in the polymerase basic protein 1 of influenza virus (PB1)(Nath, S. T. et al., 1990, Mol. Cell. Biol. 10: 4139-4145); Xenopus protein N1 (Kleinschmidt, J. A., and A. Seiter , 1988, EMBO J. 7: 1605-1614); adenovirus DNA-binding protein (DBP) (Morin, N., C. et al., 1989, Mol. Cell. Biol. 9: 4372-4380); and the yeast repressor (alpha.2 which has two nonhomologous signals, a basic NLS found at the N-terminus, as well as a signal located in the homeodomain (Hall, M. N. et al, 1984, Cell 36: 1057-1065; Hall, M. N. et al., 1990, Proc. Natl. Acad. Sci. USA 87: 6954-6958). Because each alpha.2 signal gives a different phenotype individually, Hall et al. suggest that these nonhomologous signals mediate separate steps in nuclear accumulation.

Citation of references hereinabove shall not be construed as an admission that such references are prior art to the present invention.

3. SUMMARY OF THE INVENTION

The present invention is based, in part, on the discovery and characterization by the Applicants of a gene linked to the X chromosome, *span-x*, whose novel gene product, SPAN-X, is found in X-bearing spermatazoa and in Y-bearing spermatozoa. SPAN-X is a testis-specific, haploid-expressed spermatid protein associated with the nucleus on the X chromosome, localized to nuclear vacuoles and redundant nuclear membrane. This represents the first X-specific spermatid gene to be isolated and characterized. The invention provides SPAN-X nucleic acids, polypeptides, derivatives, analogs, antibodies, ligands, host cells comprising span-x polypeptides, and the use of such compositions as a marker for spermatozoa fertility, and in gene delivery and therapy.

Specifically, the invention encompasses nucleotide sequences of *span-x* genes (human *span-x* genes and their homologs of other species). Nucleic acids hybridizable to or complementary to the foregoing nucleotide sequences are also provided.

The invention further encompasses *span-x* encoded proteins and amino acid sequences of the invention, as well as derivatives (*e.g.*, fragments) and analogs thereof. In particular, the invention encompasses SPAN-X derivatives and analogs which are functionally active, i.e., they are capable of displaying one or more functional activities associated with a full-length (wild-type) SPAN-X protein. Such functional activities include, but are not limited, to X-localization, haploid expression, antigenicity (ability to bind to an anti-SPAN-X antibody or compete with SPAN-X for binding), immunogenicity (ability to generate antibody which binds to SPAN-X). The invention further encompasses fragments (and derivatives and analogs thereof) of SPAN-X which comprise one or more domains, including, but not limited to, a nuclear localization domain, of a SPAN-X protein. In a specific embodiment, the SPAN-X protein is a human protein.

The invention further provides recombinant cells and vectors comprising *span-x* nucleic acids and recombinant SPAN-X protein. Methods of production of the SPAN-X proteins, derivatives and analogs, *e.g.*, by recombinant means, are also provided.

The invention also provides monoclonal antibodies and polyclonal antisera to SPAN-X, and SPAN-X derivatives and analogs.

The present invention also provides therapeutic and diagnostic methods and compositions based on SPAN-X proteins and nucleic acids.

In one embodiment, the invention further provides methods for the use of SPAN-X proteins as a marker competency of spermatozoa to fertilize an egg by screening for the presence of spermatozoa with or without SPAN-X.

In one embodiment, the *span-x* promoter is used for gene therapy or contraception. SPAN-X promoter sequences are used to drive spermatid-specific expression of drugs or toxins used in gene therapy in the testis. In another embodiment, the *span-x* promoter is used for contraception to drive spermatid expression of a toxin in the testis.

In one embodiment, the invention further encompasses compositions and methods for use of polynucleotides and recombinant vectors comprising the SPAN-X nuclear localization signal domain for use in gene targeting heterologous sequences to the nucleus.

The invention further provides therapeutic compounds including, but not limited to, SPAN-X proteins and analogs and derivatives (including fragments) thereof; antibodies thereto; nucleic acids encoding the SPAN-X proteins, analogs, or derivatives; and *span-x* antisense nucleic acids. In one embodiment, agonizing or antagonizing SPAN-X function can also be done to animals for veterinary sterilization purposes.

Animal models, diagnostic methods and screening methods for predisposition to disorders, and methods to identify SPAN-X agonists and antagonists, are also provided by the invention.

As used herein, underscoring or italicizing the name of a gene shall indicate the gene, in contrast to its encoded protein product which is indicated by the name of the gene in the absence of any underscoring or italicizing. For example, "*span-x*" shall mean the *span-x* gene, whereas "SPAN-X" shall indicate the protein product of the *span-x* gene.

4. DESCRIPTION OF THE FIGS.

FIG. 1. Diagram of a mature human sperm head and proximal tail. n=nucleus, m=mitochondria, rnm=redundant nuclear membrane, .a=acrosome, nv=nuclear vacuole, ax=axoneme, cd=cytoplasmic droplet, pr=posterior ring. (Modified from Holstein, A. F. and Roosen-Runge, E. C., *Atlas of human spermatogenesis* , Publisher: Berlin: Grosse, 1981).

FIG. 2. Comparison of sperm antigens recognized by the mAbs S71 and S72. Human sperm proteins were separated by either reducing or non-reducing SDS-PAGE and transferred to nitrocellulose for immunoblot analysis. Both the S71 and S72 mAbs recognized identical polypeptide bands. Under reducing conditions, major immunoreactive bands were clustered into two groups: a lower molecular weight (LMW) group of approximately 17-22 kDa and a high molecular weight (HMW) group of approximately 33-

37 kDa. Under non-reducing SDS-PAGE, only the HMW group was observed. No immunoreactivity was observed in the null ascites control. N=null ascites, AB=amido black stain.

FIG. 3. A. Comparison of cDNA of SPAN-Xa (SEQ ID NO:1) and SPAN-Xb (SEQ ID NO:3) clones. The 3' untranslated region contains a consensus polyadenylation sequence (AAUAAA; italics), a TAG termination codon (bold), and a poly A^+ tail. B. The deduced amino acid sequences of SPAN-Xa (SEQ ID NO:2) and SPAN-Xb (SEQ ID NO:4). Peptide sequences of the two overlapping consensus nuclear localization signals (PKKMKTS, shown in bold). One potential site of N-linked glycosylation was observed in the SPAN-Xb deduced peptide sequence within the six amino acid insertion (bold within insertion).

FIG. 4. A. Homology analysis of SPAN-Xa and SPAN-Xb cDNAs with four ESTs [AA382423(SEQ ID NO:5); AA412270 (SEQ ID NO:6), AA382424 (SEQ ID NO:7), and AA412605 (SEQ ID NO:8)] obtained from human testes libraries and two genomic clones [AL031078 (SEQ ID NO:9) and Z95703 (SEQ ID NO:10)]. AA412605 in this FIG. is plus strand Soares testis NHT Homo sapiens cDNA clone 730073 5'. AA412270 is minus strand Soares testis NHT Homo sapiens cDNA clone 730073 3'. AA382424 is plus strand EST95629 Testis I Homo sapiens cDNA 5' end and AA382423 is minus strand EST95628 Testis I Homo sapiens cDNA 3' end. Z95703 in this FIG. is the reverse complement of PAC 433M19 on chromosome Xq26.3-Xq27.1 . AL031078 is plus strand from genomic clone 376H23 also on the X chromosome. Small letters in the two genomic clone sequences represent intron sequences at the consensus donor/acceptor splice site junction. B. Homology analysis of SPAN-Xa and SPAN-Xb cDNAs with eight ESTs [AA382423(SEQ ID NO: 5); AA412270 (SEQ ID NO: 6), AA382424 (SEQ ID NO: 7), and AA412605 (SEQ ID NO: 8); AI143898 (SEQ ID NO: 24); AI962751 (SEQ ID NO: 25); AI808260 (SEQ ID NO: 26); and AI208372 (SEQ ID NO: 27)] obtained from human testes libraries. C. The diagram shows the intron-exon organization of the polymorphic SPAN-X gene. Exon I differs by the 18 bp insertion observed in the SPAN-Xb cDNA and the genomic sequence AL031078 (hatched box).

FIG. 5. Analysis of SPAN-Xa and SPAN-Xb deduced peptide sequence. On the top line, the basic and acidic amino acid residues are presented above and below the line, respectively. The hydrophobic and hydrophilic amino acids are as indicated. The Kyte-Doolittle plot of hydrophobicity is represented by the second line demonstrating that both

peptide sequences are largely hydrophilic and do not contain potential transmembrane domains.

FIG. 6. Northern analysis of SPAN-X transcript demonstrating testis-specific expression. Poly(A)+ RNA from a panel of human tissues was examined by Northern blotting with DIG-labeled SPAN-Xa cDNA probe. A transcript of 0.61 kb was observed exclusively in human testis. No transcripts were identified in other human tissues including peripheral blood leukocytes, colon, small intestine, ovary, prostate, thymus, and spleen. β-actin signal of approximately 2.0 kb was observed in all tissues.

FIG. 7. *In situ* hybridization of human testis sections. Sections hybridized with a tritium-labeled SPAN-Xa antisense riboprobe (complementary to endogenous tissue mRNA) demonstrate silver grains primarily in the most luminal cell types corresponding to round and elongating spermatids (A. B). Few silver grains were observed over more mature elongated spermatids (C). Only background staining was observed on spermatocytes and other cell types and with the control sense probe (D). bm=basement membrane. L=lumen, c=spermatocytes, r=round spermatids, e=elongating spermatids.

FIG 8. Western analysis of IMAC-purified recSPAN-X protein stained with amido black staining or S71 mAb. Positive immunoreactivity was observed with the major band in the purified recombinant material.

FIG. 9. Western analysis of human sperm extracts stained with amido black (AB) or immunostained with pre- or post-immune sera (PI and pAb, respectively) generated in guinea pig against the recSPAN-X or with the S71 mAb (mAb). On Western blots of reduced human sperm extracts, anti-recSPAN-X pAbs reacted with a series of polymorphic bands between 15-20 kDa identical to the LMW group immunoreactive with the S71 mAb. On Western blots of non-reduced sperm extracts, anti-recSPAN-X antibodies recognized two series of polymorphic bands between 15-20 kDa and between 27-30 kDa. Pre-immune sera showed only background reactivity in all samples.

FIG. 10. Western blot immunostained with anti-SPAN-X antiserum showing the solubility properties of SPAN-X. The supernatant (S) fluid and pellet (P) fractions of sperm extracted with 0.5% CHAPS, with 0.5% CHAPS and 600 mM KCl, with 0.5% CHAPS, 600 mM KCl, and 2mM DTT, or with 1.0% SDS are presented. SPAN-X was solubilized with

denaturing detergent but remained in the insoluble pellet fractions following extraction with zwitterionic detergent, high salt and reducing agents.

FIG. 11. A. Two-dimensional electrophoresis of human sperm extracts from four donors stained with silver (a, a') or immunostained with anti-recSPAN-X (b, b'). a' and b' are enlarged images of the boxed area in a. The immunoreactive spots (b') correspond to yellow stained proteins on the silver-stained gel (a'). A group of 19 spots was immunostained with the pAb.

FIG. 12. Indirect immunofluorescent staining of fixed, permeabilized swim-up spermatozoa with the anti-recSPAN-X immune serum or pre-immune serum. Matched DIC (top image), epifluorescence (FITC; middle image) and dual DIC/FITC images (bottom image). Immunofluorescence is observed in association with the nuclear craters (large arrows), in the cytoplasmic droplet at the posterior of the sperm head (small arrow), or in both. Notice that not all spermatozoa in the field demonstrate immunofluorescent staining (i.e. dashed circle). No staining is observed in the remainder of the head or tail of the sperm or when the sperm are stained with pre-immune sera.

FIG. 13. Indirect immunofluorescent staining of human ejaculated spermatozoa showing distinct SPAN-X patterns of labeling. Matched DIC (left), epifluorescence (center) and dual DIC/fluorescence images (right). Immunofluorescence is observed in association with small and large nuclear craters (a, b, respectively), in small and large cytoplasmic droplets at the posterior of the sperm head (c, d) or in both nuclear craters and cytoplasmic droplets (e, f).

FIG. 14. Electron micrographs of human spermatozoa post-embedding immunolabeled with anti-recSPAN-X antibodies. Intense gold labeling is present on granular material within nuclear vacuoles (A, B) that often appeared to be associated with membrane structures. In addition, specific staining of the redundant nuclear membrane was observed in the cytoplasmic droplet at the base of the sperm head (A, C). a=acrosome, es=equatorial segment of the acrosome, n=nucleus, nv=nuclear vacuole, rnm=redundant nuclear membrane, bp=basal plate, pr=posterior ring boundary.

FIG. 15. Northern analysis was performed on Northern blots containing testicular mRNA from four species, human, cynomologus macaque, rat and mouse.

FIG. 16. Indirect immunofluorescence analysis of human and chimpanzee testicular tissue to examine the localization of SPAN-X during spermatogenesis.

5. DETAILED DESCRIPTION OF THE INVENTION

The present invention provides "*span-x*"(sperm protein associated with the nucleus on the X chromosome) polynucleotides, polypeptides, and derivatives and analogs thereof. Production of recombinant SPAN-X proteins, fragments comprising functional domains, derivatives, antibodies, and ligands are also provided. The invention further provides compositions, kits, and methods for their use to separate fertilization competent spermatozoa from fertilization incompetent spermatozoa. The invention further provides therapeutic compositions and methods for the use of *span-x* polynucleotides for targeted heterologous gene expression.

The invention further encompasses the use of nucleotides encoding SPAN-X proteins and peptides, as well as antibodies to SPAN-X (which can, for example, act as agonists or antagonists), ligands that bind to SPAN-X or modulate the function, activity or expression of SPAN-X. In addition, regulatory nucleotides and nucleotides encoding SPAN-X polypeptides or one or more functional domains of SPAN-X or fragments thereof, are effective in gene therapy, or for delivery of heterologous gene products to a cellular or subcellular locale.

In particular, embodiments of the invention described in the subsections below encompasses SPAN-X, polypeptides or peptides corresponding to functional domains of SPAN-X (*e.g.*, the nuclear localization domain, a nuclear-scaffold binding domain, and a ligand-binding domain), mutated, truncated or deleted (*e.g.* with one or more functional domains or portions thereof deleted), SPAN-X fusion proteins, nucleotide sequences encoding such products, and host cell expression systems that can produce such SPAN-X products.

The invention also encompasses antibodies, including anti-idiotypic antibodies, antagonists and agonists, as well as compounds or nucleotide constructs that inhibit expression of the SPAN-X gene (transcription factor inhibitors, antisense and ribozyme molecules, or gene or regulatory sequence replacement constructs), or promote expression of SPAN-X (*e.g.*, expression constructs in which SPAN-X coding sequences are operatively associated with expression control elements such as promoters, promoter/enhancers, etc.). The invention also provides host cells and animals genetically engineered to express the human (or mutants thereof) SPAN-X regulatory or protein coding sequences, or to inhibit or "knock-out" expression of the animal's endogenous SPAN-X.

The SPAN-X products and fusion protein products, (*i.e.*, fusions of the proteins or a domain of the protein, *e.g.*, the nuclear localization domain), antibodies and anti-idiotypic antibodies (including Fab fragments), modulators and ligands can be used to separate fertilization competent from incompetent spermatozoa by screening for the presence or absence of SPAN-X.

The present invention provides methods of screening for agents, small molecules, or proteins that interact with SPAN-X. The invention encompasses both *in vivo* and *in vitro* assays to screen small molecules, compounds, recombinant proteins, peptides, nucleic acids, antibodies etc. which bind to or modulate the activity of SPAN-X and are thus useful as diagnostic markers for spermatozoa fertility by screening for the presence of presence of spermatozoa with altered fertility in ejaculated spermatozoa.

The *span-x* polynucleotides (*i.e.*, regulatory regions of the *span-x* gene) and fusion protein products, (*i.e.*, fusions of the SPAN-X proteins or a domain of the protein, *e.g.*, the nuclear localization domain to another heterologous polypeptide), antibodies (including, but without limitation, anti-idiotypic antibodies, and Fab fragments), modulators and ligands can be used for drug delivery or gene therapy. Thus, the invention also encompasses pharmaceutical formulations and methods for contraception and treating infertility and cancer.

Various aspects of the invention are described in greater detail in the subsections below.

5.1. *SPAN-X* NUCLEIC ACIDS

The invention provides the nucleotide sequences of *span-x* nucleic acids which were identified by their presence in testis specific cDNA libraries. Nucleic acid sequences of the identified *span-x* genes are described herein. As used herein, "a *span-x* nucleic acid" refers to:

(a) a nucleic acid molecule containing the nucleotide sequence of *span-xa* (SEQ ID NO:1) or *span-xb* (SEQ ID NO:3), shown in FIG. 3A;

(b) any nucleotide sequence that encodes a polypeptide containing the amino acid sequence of SPAN-Xa (SEQ ID NO:2) or the amino acid sequence of SPAN-Xb (SEQ ID NO:4) shown in FIG. 3B;

(c) any nucleotide sequence that hybridizes to the complement of the DNA sequences that encode any of the amino acid sequences of SPAN-Xa (SEQ ID NO: 2) or SPAN-Xb (SEQ ID NO: 4) under highly stringent conditions, *e.g.*, hybridization to filter-bound DNA in 0.5 M $NaHPO_4$, 7% sodium dodecyl sulfate (SDS), 1 mM EDTA at 65°C, and washing in 0.1xSSC/0.1% SDS at 68°C (Ausubel

F.M. *et al.*, eds., 1989, Current Protocols in Molecular Biology, Vol. I, Green Publishing Associates, Inc., and John Wiley & sons, Inc., New York, at p. 2.10.3); and/or

(d) any DNA sequence that hybridizes to the complement of the DNA sequences that encode any of the amino acid sequences in SPAN-Xa (SEQ ID NO: 2) or SPAN-Xb (SEQ ID NO: 4), under less stringent conditions, such as moderately stringent conditions, *e.g.*, washing in 0.2xSSC/0.1% SDS at 42°C (Ausubel *et al.*, 1989, *supra*), and encodes a gene product functionally equivalent to a *span-x* gene product,

provided that such nucleotide sequence does not consist of the nucleotide sequence of one or more of the following previously described ESTs: Genbank Accession No. AA382423 (SEQ ID NO:5); Genbank Accession No. AA412270 (SEQ ID NO:6); Genbank Accession No. AA382424 (SEQ ID NO:7); and Genbank Accession No.AA412605 (SEQ ID NO:8) obtained from human testes libraries and shown in FIG. 4, or the entire sequence or a subfragment thereof of Genbank Accession No. AL031078 (SEQ ID NO:9) and Genbank Accession No. Z95703 (SEQ ID NO:10), obtained from genomic clones of the human X chromosome, shown in FIG. 4.

The invention also includes nucleic acid molecules derived from mammalian nucleic acids, preferably DNA molecules, that hybridize to, and are therefore the complements of, the nucleotide sequences (a) through (d), in the preceding paragraph. Such hybridization conditions may be highly stringent or less highly stringent, as described above. In instances wherein the nucleic acid molecules are deoxyoligonucleotides ("oligos"), highly stringent conditions may refer, *e.g.*, to washing in 6xSSC/0.05% sodium pyrophosphate at 37°C (for 14-base oligos), 48°C (for 17-base oligos), 55°C (for 20-base oligos), and 60°C (for 23-base oligos). These nucleic acid molecules may encode or act as *span-x* nucleic acid antisense molecules, useful, for example, in *span-x* gene regulation (for and/or as antisense primers in amplification reactions of *span-x* nucleic acid sequences). With respect to *span-x* gene regulation, such techniques can be used to regulate, for example, *a span-x*-regulated pathway, in order to block cell proliferation associated with cancer. Further, such sequences may be used as part of ribozyme and/or triple helix sequences, also useful for *span-x* gene regulation. Still further, such molecules may be used as components of diagnostic methods whereby, for example, the presence of a particular *span-x* allele responsible for causing *a span-x* related disorder, *e.g.*, fertility or proliferative disorders such as infertility or cancer, may be detected.

The invention further includes fragments of any of the nucleotide sequences disclosed herein.

In a specific embodiment, the fragment of a *span-x* nucleic acid encodes a *span-x* nuclear localization signal sequence.

As used herein, a *span-x* nuclear localization signal sequence includes:

(a) the nucleic acid molecule (SEQ ID NO:18) which encodes the consensus nuclear localization signal at amino acids 37-43 (PAPKKMK)(SEQ ID NO:15), the nucleic acid molecule (SEQ ID NO:19) that encodes the consensus nuclear localization signal at amino acids 39-45 (PKKMKTS)(SEQ ID NO:16), or the nucleic acid molecule (SEQ ID NO:20) that encodes the overlapping or embedded consensus nuclear localization signal sequence spanning amino acids 37-45 (PAPKKMKTS)(SEQ ID NO:17);

(b) any DNA sequence that encodes the consensus nuclear localization signal at amino acids 37-43 (PAPKKMK)(SEQ ID NO:15), or the consensus nuclear localization signal at amino acids 39-45 (PKKMKTS)(SEQ ID NO: 16), or the overlapping or embedded consensus nuclear localization signal sequence spanning amino acids 37-45 (PAPKKMKTS)(SEQ ID NO:17);

(c) any DNA sequence that hybridizes to the complement of the DNA sequences that encode the nuclear localization signal sequences described in (a) or (b) under highly stringent conditions, *e.g.*, hybridization to filter-bound DNA in 0.5 M $NaHPO_4$, 7% sodium dodecyl sulfate (SDS), 1 mM EDTA at 65°C, and washing in 0.1xSSC/0.1% SDS at 68°C (Ausubel F.M. *et al.*, eds., 1989, Current Protocols in Molecular Biology, Vol. I, Green Publishing Associates, Inc., and John Wiley & sons, Inc., New York, at p. 2.10.3); and/or

(d) any DNA sequence that hybridizes to the complement of the DNA sequences that encode the consensus nuclear localization signal sequences described in (a) or (b) under less stringent conditions, such as moderately stringent conditions, *e.g.*, washing in 0.2xSSC/0.1% SDS at 42°C (Ausubel *et al.*, 1989, *supra*), and encodes a gene product functionally equivalent to a *span-x* nuclear localization signal.

The invention encompasses the *span-x* nuclear localization signal sequences, in isolated or purified form, as well as compositions containing such nuclear localization signal sequence operatively associated with a nucleic acid encoding a protein or polypeptide heterologous to SPAN-X.

Span-x sequences of the present invention are derived from a eukaryotic genome, preferably a mammalian genome, and more preferably a human genome. In a specific

embodiment, the nucleotides of the present invention encompass any DNA sequence derived from a mammalian genome which hybridizes under highly stringent conditions to SEQ ID NO: 1 or SEQ ID NO: 3. In a specific embodiment, the nucleotides of the present invention encompass any DNA sequence derived from a mammalian genome which hybridize under highly stringent conditions to SEQ ID NO: 1 or SEQ ID NO: 3 and encodes a gene product which is expressed in X-and Y-bearing spermatids, and contains a nuclear localization domain.

When referring to a nucleic acid which encodes a given amino acid sequence, it is to be understood that the nucleic acid need not only be a cDNA molecule, but can also, for example, refer to a gDNA sequence from which an mRNA species is transcribed that is processed to encode the given amino acid sequence.

The invention further includes regulator nucleic acids of the *span-x* gene. The genomic sequence of the *span-x* gene contains regulatory sequences in the non-coding 5'-flanking region. The 5'-regulatory sequences of the *span-x* gene comprise the polynucleotide sequences located between the nucleotide in position -2000 and the nucleotide in position +103 of the nucleotide sequence of SEQ ID NO 21 or SEQ ID NO 22, or more preferably between positions -2074 and +103 of SEQ ID NO. 21 or SEQ ID NO: 22.

The invention also encompasses:

(a) vectors that contain any of the foregoing *span-x* coding sequences and/or their complements (*i.e.*, antisense);

(b) expression vectors that contain any of the foregoing *span-x* coding sequences operatively associated with a regulatory element that directs the expression of the coding sequences; and

(c) genetically engineered host cells that contain any of the foregoing *span-x* coding sequences operatively associated with a regulatory element that directs the expression of the coding sequences in the host cell.

As used herein, regulatory elements include but are not limited to inducible and non-inducible promoters, enhancers, operators and other elements known to those skilled in the art that drive and regulate expression. Such regulatory elements include but are not limited to the cytomegalovirus hCMV immediate early gene, the early or late promoters of SV40 adenovirus, the *lac* system, the *trp* system, the *TAC* system, the *TRC* system, the major operator and promoter regions of phage A, the control regions of fd coat protein, the promoter for 3-phosphoglycerate kinase, the promoters of acid phosphatase, and the promoters of the yeast α-mating factors.

In one embodiment, the *span-x* nucleic acid sequences of the invention are mammalian nucleic acid sequences, with human sequences being preferred.

In yet another embodiment, the *span-x* nucleic acid sequences of the invention are nucleic acid sequences encoding *span-x* gene products containing polypeptide portions corresponding to (that is, polypeptide portions exhibiting amino acid sequence similarity to) the amino acid sequences depicted in FIG. 3B, wherein the corresponding portion exhibits greater than about 50% amino acid identity with the depicted sequence, averaged across the *span-x* gene product's entire length.

In specific embodiments, SPAN-X encoding nucleic acids comprise the cDNA sequences of SEQ ID NOs:1 or 3 nucleotide sequence of FIG. 3A or the coding regions thereof, or nucleic acids encoding a SPAN-X protein (*e.g.*, a protein having the sequence of SEQ ID NOs:2 or 4 as shown in FIG. 3B). The invention provides isolated or purified nucleic acids consisting of at least 8 nucleotides (*i.e.*, a hybridizable portion) of *a span-x* nucleic acid sequence; in other embodiments, the nucleic acids consist of at least 25 (continuous) nucleotides, 50 nucleotides, 100 nucleotides, 150 nucleotides, 200 nucleotides, or 469 contiguous nucleotides of a SPAN-X sequence, or a full-length SPAN-X coding sequence. For example, in one embodiment the invention provides isolated or purified nucleic acids consisting of nucleotides 1-108, 201-250, 251-300, 301-350, 351-400, 401-450, or 451-469 of SEQ ID NO: 1. For example, in one embodiment the invention provides purified nucleic acids consisting of nucleotides 1-50, 51-100, 101-150, 151-200, 201-250, 251-300, 301-350, 351-400, 401-450, or 451-469 of SEQ ID NO: 3. In another embodiment, the nucleic acids are smaller than 35, 200 or 500 nucleotides in length. Nucleic acids can be single or double stranded. The invention also provides nucleic acids hybridizable to or complementary to the foregoing sequences. In specific aspects, nucleic acids are provided which comprise a sequence complementary to at least 10, 25, 50, 100, or 200 nucleotides or the entire coding region of a *span-x* gene. As will be understood by those skilled in the art, the invention also encompasses those genomic DNA sequences which give rise to the cDNA sequences of SEQ ID NOs: 1 and 3 described above.

In addition to the human *span-x* nucleic acid sequences disclosed in SEQ ID NOS:1 or 3 shown in FIG. 3A, other *span-x* nucleic acid sequences can be identified and readily isolated, without undue experimentation, by molecular biological techniques well known in the art, used in conjunction with the *span-x* nucleic acid sequences disclosed herein. These other sequences are encompassed by the present invention. For example, additional human *span-x* nucleic acid sequences at the same or at different genetic loci as those disclosed in SEQ ID NOs:1 or 3 can be isolated readily. There can exist, for example, genes at other genetic or physical loci within the human genome that encode proteins that have extensive

homology to one or more domains of the *span-x* gene products and that encode gene products functionally equivalent to a *span-x* gene product. Further, homologous *span-x* nucleic acid sequences present in other species can be identified and isolated readily.

With respect to identification and isolation of *span-x* nucleic acid sequences present at the same genetic or physical locus as those sequences disclosed in SEQ ID NOs.1 or 3, such sequences can, for example, be obtained readily by utilizing standard sequencing and bacterial artificial chromosome (BAC) and P1 artificial chromosome (PAC) technologies.

With respect to the cloning of a *span-x* gene or nucleic acid homologue in human or other species (*e.g.*, mouse), the isolated *span-x* nucleic acid sequences disclosed herein may be labeled and used to screen a cDNA library constructed from mRNA obtained from appropriate cells or tissues (*e.g.*, testes) derived from the organism (*e.g.*, mouse) of interest. The hybridization conditions used should be of a lower stringency when the cDNA library is derived from an organism different from the type of organism from which the labeled sequence was derived.

Alternatively, the labeled fragment may be used to screen a genomic library derived from the organism of interest, again, using appropriately stringent conditions. Low stringency conditions are well known to those of skill in the art, and will vary predictably depending on the specific organisms from which the library and the labeled sequences are derived. For guidance regarding such conditions see, for example, Sambrook, *et al.*, 1989, Molecular Cloning, A Laboratory Manual, Second Edition, Cold Spring Harbor Press, N.Y.; and Ausubel, *et al.*, *supra*. Further, a *span-x* gene homologue may be isolated from, for example, human nucleic acid, by performing PCR using two degenerate oligonucleotide primer pools designed on the basis of amino acid sequences within any *span-x* gene product disclosed herein.

The PCR product may be subcloned and sequenced to ensure that the amplified sequences represent the sequences of a *span-x* gene nucleic acid sequence. The PCR fragment may then be used to isolate a full length cDNA clone by a variety of methods. For example, the amplified fragment may be labeled and used to screen a bacteriophage cDNA library. Alternatively, the labeled fragment may be used to isolate genomic clones via the screening of a genomic library.

PCR technology may also be utilized to isolate full length cDNA sequences. For example, RNA may be isolated, following standard procedures, from an appropriate cellular or tissue source (*i.e.*, testis). A reverse transcription reaction may be performed on the RNA using an oligonucleotide primer specific for the most 5' end of the amplified fragment for the priming of first strand synthesis. The resulting RNA/DNA hybrid may then be "tailed" with guanines using a standard terminal transferase reaction, the hybrid may be

digested with RNAase H, and second strand synthesis may then be primed with a poly-C primer. Thus, cDNA sequences upstream of the amplified fragment may easily be isolated. For a review of cloning strategies that may be used, see *e.g.*, Sambrook *et al.*, *supra*.

Span-x nucleic acid sequences may additionally be used to identify mutant *span-x* gene alleles. Such mutant alleles may be isolated from individuals either known or proposed to have a genotype that contributes to the symptoms of a *span-x* gene disorder, such as fertility disorders, for example. Such alleles are encompassed by the present invention.

Span-x alleles may be identified by single strand conformational polymorphism (SSCP) mutation detection techniques, Southern blot, and/or PCR amplification techniques. Primers can routinely be designed to amplify overlapping regions of the whole *span-x* sequence including the promoter region. In one embodiment, primers are designed to cover the exon-intron boundaries such that, first, coding regions can be scanned for mutations. Genomic DNA isolated from lymphocytes of normal and affected individuals is used as PCR template. PCR products from normal and affected individuals are compared, either by single strand conformational polymorphism (SSCP) mutation detection techniques and/or by sequencing. SSCP analysis can be performed as follows: 100 ng of genomic DNA is amplified in a 10 μl reaction, adding 10 pmols of each primer, 0.5 U of Taq DNA polymerase (Promega), 1 μCi of α-[^{32}P]dCTP (NEN; specific activity, 3000 Ci/mmol), in 2.5 μM dNTPs (Pharmacia), 10 mM Tris-HCl (pH 8.8), 50 mM KCl, 1 mM MgCl2, 0.01% gelatin, final concentration. Thirty cycles of denaturation (94°C), annealing (56°C to 64°C, depending on primer melting temperature), and extension (72°C) is carried out in a thermal-cycler (MJ Research, Boston, MA, USA), followed by a 7 min final extension at 72°C. Two microliters of the reaction mixture is diluted in 0.1% SDS, 10 mM EDTA and then mixed 1: 1 with a sequencing stop solution containing 20 mM NaOH. Samples are heated at 95°C for 5 min, chilled on ice for 3 min and then 3 μl will be loaded onto a 6% acrylamide/TBE gel containing 5% (v/v) glycerol. Gels are run at 8 W for 12-15 h at room temperature. Autoradiography is performed by exposure to film at -70°C with intensifying screes for different periods of time. The mutations responsible for the loss or alteration of function of the mutant *span-x* gene product can then be ascertained.

Alternatively, a cDNA of a mutant *span-x* gene may be isolated, for example, using PCR. In this case, the first cDNA strand may be synthesized by hybridizing an oligo-dT oligonucleotide to mRNA isolated from tissue known or suspected to be expressed in an individual putatively carrying the mutant *span-x* allele, and by extending the new strand with reverse transcriptase. The second strand of the cDNA is then synthesized using an oligonucleotide that hybridizes specifically to the 5' end of the normal gene. Using these

two primers, the product is then amplified via PCR, cloned into a suitable vector, and subjected to DNA sequence analysis through methods well known to those of skill in the art. By comparing the DNA sequence of the mutant *span-x* allele to that of the normal *span-x* allele, the mutation(s) responsible for the loss or alteration of function of the mutant *span-x* gene product can be ascertained.

Alternatively, a genomic library can be constructed using DNA obtained from an individual suspected of or known to carry a mutant *span-x* allele, or a cDNA library can be constructed using RNA from a tissue known, or suspected, to express a mutant *span-x* allele. An unimpaired *span-x* gene or any suitable fragment thereof may then be labeled and used as a probe to identify the corresponding mutant *span-x* allele in such libraries. Clones containing the mutant *span-x* nucleic acid sequences may then be purified and subjected to sequence analysis according to methods well known to those of skill in the art.

Additionally, an expression library can be constructed utilizing cDNA synthesized from, for example, RNA isolated from a tissue known, or suspected, to express a mutant *span-x* allele in an individual suspected of or known to carry such a mutant allele. In this manner, gene products made by the putatively mutant tissue may be expressed and screened using standard antibody screening techniques in conjunction with antibodies raised against the normal *span-x* gene product, as described, below, in Section 5.3. (For screening techniques, see, for example, Harlow and Lane, eds., 1988, "Antibodies: A Laboratory Manual", Cold Spring Harbor Press, Cold Spring Harbor.)

Nucleic acids encoding derivatives and analogs of SPAN-X proteins, and SPAN-X antisense nucleic acids can be isolated by the methods recited above. As used herein, a "nucleic acid encoding a fragment or portion of a SPAN-X protein" shall be construed as referring to a nucleic acid encoding only the recited fragment or portion of the SPAN-X and not the other contiguous portions of the SPAN-X protein as a continuous sequence.

Fragments of *span-x* nucleic acids comprising regions conserved between (*i.e.*, with homology to) other *span-x* nucleic acids, of the same or different species, are also provided. Nucleic acids encoding one or more SPAN-X domains can be isolated by the methods recited above.

In cases where a *span-x* mutation results in an expressed gene product with altered function (*e.g.*, as a result of a missense or a frameshift mutation), a polyclonal set of anti-*span-x* gene product antibodies are likely to cross-react with the mutant *span-x* gene product. Library clones detected via their reaction with such labeled antibodies can be purified and subjected to sequence analysis according to methods well known to those of skill in the art.

5.2. SPAN-X PROTEINS AND POLYPEPTIDES

The invention further provides *span-x* encoded proteins and amino acid sequences, as well as derivatives (*e.g.*, fragments) and analogs thereof. In particular, the invention provides SPAN-X derivatives and analogs which are functionally active, i.e., they are capable of displaying one or more functional activities associated with a full-length (wild-type) SPAN-X protein. Such functional activities include, but are not limited to, spermatozoa markers, haploid expression, antigenicity (ability to bind to an anti-SPAN-X antibody or compete with SPAN-X for binding), immunogenicity (ability to generate antibody which binds to SPAN-X). The invention further provides fragments (and derivatives and analogs thereof) of SPAN-X which comprise one or more domains of a SPAN-X protein. In a specific embodiment, the SPAN-X protein is a human protein.

The amino acid sequences depicted in FIG. 3B (SEQ ID NOs:2 and 4) represent *span-x* gene products. The *span-x* gene product, sometimes referred to herein as a "SPAN-X", or "SPAN-Xa" and "SPAN-Xb" includes those products encoded by the *span-x* nucleic acid sequences described in Section 5.1, above. In accordance with the present invention, the nucleic acid sequences encoding the *span-x* gene products are derived from eukaryotic genomes, including mammalian genomes. In a preferred embodiment the nucleic acid sequences encoding the *span-x* products are derived from the human genome.

SPAN-X proteins, polypeptides and peptide fragments thereof, can be prepared for a variety of uses. For example, such molecules can be used for the generation of antibodies, for use in diagnostic and therapeutic assays, or for the identification of other cellular, membrane-associated, or extracellular gene products expressed in X-bearing sperm or during development of X-bearing spermatazoa .

In addition, *span-x* products of the present invention may include proteins that represent functionally equivalent (see Section 5.1 for a definition) products. Functionally equivalent *span-x* products may contain deletions, including internal deletions, additions, including additions yielding fusion proteins, or substitutions of amino acid residues within and/or adjacent to the amino acid sequence encoded by the *span-x* nucleic acid sequences described, above, in Section 5.1, but that result in a "silent" change, in that the change produces a functionally equivalent *span-x* gene product. Amino acid substitutions may be made on the basis of similarity in polarity, charge, solubility, hydrophobicity, hydrophilicity, and/or the amphipathic nature of the residues involved. For example, nonpolar (hydrophobic) amino acids include alanine, leucine, isoleucine, valine, proline, phenylalanine, tryptophan, and methionine; polar neutral amino acids include glycine, serine, threonine, cysteine, tyrosine, asparagine, and glutamine; positively charged (basic)

amino acids include arginine, lysine, and histidine; and negatively charged (acidic) amino acids include aspartic acid and glutamic acid.

Specific functional variants of SPAN-X proteins encompassed by the present invention include immunoreactive proteins with the following molecular weights (MW) in kiloDaltons, and isoloelectric points (pI), obtainable from human sperm preparations: a protein of MW 22.8 and pI 4.99; a protein of MW 21.9 and pI 5.02; a protein of MW 20.9 and pI 5.00; protein of MW 23.2 and pI 5.03; a protein of MW 22.6 and pI 5.06; a protein of MW 21.3 and pI 5.09; a protein of MW 20.6 and pI 5.06; a protein of MW 23.5 and pI 5.09; a protein of MW 23.0 and pI 6.30; a protein of MW 22.4 and pI 6.24; a protein of MW 22.1 and pI 5.20; a protein of MW 21.5 and pI 5.24; a protein of MW 21.6 and pI 5.30; a protein of MW 20.1 and pI 5.26; a protein of MW 20.3 and pI 6.23; a protein of MW 20.2 and pI 5.28; a protein of MW 21.6 and pI 5.54; a protein of MW 21.3 and pI 5.53; and a protein of MW 19.8 and pI 5.44.

Alternatively, where alteration of function is desired, deletion or non-conservative alterations can be engineered to produce altered *span-x* gene products. Such alterations can, for example, alter one or more of the biological functions of the *span-x* product. Further, such alterations can be selected so as to generate *span-x* products that are better suited for expression, scale up, etc. in the host cells chosen. For example, cysteine residues can be deleted or substituted with another amino acid residue in order to eliminate disulfide bridges.

5.2.1. RECOMBINANT SPAN-X PROTEINS, VECTORS, AND CELLS

The *span-x* products, peptide fragments thereof and fusion proteins thereof, may be produced by recombinant DNA technology using techniques well known in the art. For example, the production and purification of recombinant SPAN-X proteins are exemplified by the recombinant production of recSPAN-Xa and recSPAN-Xb in Section 6. Thus, methods for preparing the SPAN-X polypeptides, peptides, fusion peptide and fusion polypeptides of the invention by expressing nucleic acid containing *span-x* nucleic acid sequences are described herein. Methods that are well known to those skilled in the art can be used to construct expression vectors containing SPAN-X coding sequences and appropriate transcriptional and translational control signals. These methods include, for example, *in vitro* recombinant DNA techniques, synthetic techniques, and *in vivo* genetic recombination. See, for example, the techniques described in Sambrook, *et al.*, *supra*, and Ausubel, *et al.*, *supra*. Alternatively, RNA capable of encoding *span-x* gene product sequences may be chemically synthesized using, for example, synthesizers. See, for

example, the techniques described in "Oligonucleotide Synthesis", 1984, Gait, ed., IRL Press, Oxford.

A variety of host-expression vector systems may be utilized to express the SPAN-X coding sequences of the invention. Such host-expression systems represent vehicles by which the coding sequences of interest may be produced and subsequently purified, but also represent cells that may, when transformed or transfected with the appropriate nucleotide coding sequences, exhibit the SPAN-X product of the invention *in situ*. These include but are not limited to microorganisms such as bacteria (*e.g.*, *E. coli*, *B. subtilis*) transformed with recombinant bacteriophage DNA, plasmid DNA or cosmid DNA expression vectors containing *span-x* gene product coding sequences; yeast (*e.g.*, *Saccharomyces*, *Pichia*) transformed with recombinant yeast expression vectors containing the *span-x* gene product coding sequences; insect cell systems infected with recombinant virus expression vectors (*e.g.*, baculovirus) containing the *span-x* gene product coding sequences; plant cell systems infected with recombinant virus expression vectors (*e.g.*, cauliflower mosaic virus, CaMV; tobacco mosaic virus, TMV) or transformed with recombinant plasmid expression vectors (*e.g.*, Ti plasmid) containing *span-x* gene product coding sequences; or mammalian cell systems (*e.g.*, COS, CHO, BHK, 293, 3T3) harboring recombinant expression constructs containing promoters derived from the genome of mammalian cells (*e.g.*, metallothionein promoter) or from mammalian viruses (*e.g.*, the adenovirus late promoter; the vaccinia virus 7.5K promoter).

In bacterial systems, a number of expression vectors may be advantageously selected depending upon the use intended for the SPAN-X product being expressed. For example, when a large quantity of such a protein is to be produced, for the generation of pharmaceutical compositions of SPAN-X protein or for raising antibodies to SPAN-X protein, for example, vectors that direct the expression of high levels of fusion protein products that are readily purified may be desirable. Such vectors include, but are not limited, to the *E. coli* expression vectors of the pET series (Novagen), in which the *span-x* nucleic acid may be expressed under the control of the T7 polymerase promoter; the pUR278 (Ruther *et al.*, 1983, EMBO J. 2, 1791), in which the *span-x* nucleic coding sequence may be ligated individually into the vector in frame with the *lac Z* coding region so that a fusion protein is produced; pIN vectors (Inouye and Inouye, 1985, Nucleic Acids Res. 13, 3101-3109; Van Heeke and Schuster, 1989, J. Biol. Chem. 264, 5503-5509); and the like. pGEX vectors may also be used to express foreign polypeptides as fusion proteins with glutathione S-transferase (GST). In general, such fusion proteins are soluble and can easily be purified from lysed cells by adsorption to glutathione-agarose beads followed by elution in the presence of free glutathione. The pGEX vectors are designed to include

thrombin or factor Xa protease cleavage sites so that the cloned target gene product can be released from the GST moiety.

In an insect system, *Autographa californica*, nuclear polyhedrosis virus (AcNPV) is used as a vector to express foreign genes. The virus grows in *Spodoptera frugiperda* cells. The *span-x* nucleic acid coding sequence may be cloned individually into non-essential regions (for example the polyhedrin gene) of the virus and placed under control of an AcNPV promoter (for example the polyhedrin promoter). Successful insertion of *span-x* gene coding sequence will result in inactivation of the polyhedrin gene and production of non-occluded recombinant virus (*i.e.*, virus lacking the proteinaceous coat coded for by the polyhedrin gene). These recombinant viruses are then used to infect *Spodoptera frugiperda* cells in which the inserted gene is expressed. (*e.g.*, see Smith, *et al.*, 1983, J. Virol. 46, 584; Smith, U.S. 4,215,051).

In mammalian host cells, a number of viral-based expression systems may be utilized. In cases where an adenovirus is used as an expression vector, the *span-x* nucleic acid coding sequence of interest may be ligated to an adenovirus transcription/translation control complex, *e.g.*, the major late promoter and tripartite leader sequence. This chimeric gene may then be inserted in the adenovirus genome by *in vitro* or *in vivo* recombination. Insertion in a non-essential region of the viral genome (*e.g.*, region E1 or E3) will result in a recombinant virus that is viable and capable of expressing *span-x* gene product in infected hosts. (*e.g.*, See Logan and Shenk, 1984, Proc. Natl. Acad. Sci. USA 81, 3655-3659). Specific initiation signals may also be required for efficient translation of inserted *span-x* nucleic acid product coding sequences. These signals include the ATG initiation codon and adjacent sequences. In cases where an entire *span-x* gene, including its own initiation codon and adjacent sequences, is inserted into the appropriate expression vector, no additional translational control signals may be needed. However, in cases where only a portion of the *span-x* nucleic acid coding sequence is inserted, exogenous translational control signals, including, perhaps, the ATG initiation codon, must be provided. Furthermore, the initiation codon must be in phase with the reading frame of the desired coding sequence to ensure translation of the entire insert. These exogenous translational control signals and initiation codons can be of a variety of origins, both natural and synthetic. The efficiency of expression may be enhanced by the inclusion of appropriate transcription enhancer elements, transcription terminators, etc. (see Bittner, *et al.*, 1987, Methods in Enzymol. 153, 516-544).

In addition, a host cell strain may be chosen that modulates the expression of the inserted sequences, or modifies and processes the gene product in the specific fashion desired. Such modifications (*e.g.*, glycosylation) and processing (*e.g.*, cleavage) of protein

products may be important for the function of the protein. Different host cells have characteristic and specific mechanisms for the post-translational processing and modification of proteins and gene products. Appropriate cell lines or host systems can be chosen to ensure the correct modification and processing of the foreign protein expressed. To this end, eukaryotic host cells that possess the cellular machinery for proper processing of the primary transcript, glycosylation, and phosphorylation of the gene product may be used. Such mammalian host cells include, but are not limited to, CHO, VERO, BHK, HeLa, COS, MDCK, 293, 3T3, and WI38.

For long-term, high-yield production of recombinant proteins, stable expression is preferred. For example, cell lines that stably express the SPAN-X product may be engineered. Rather than using expression vectors that contain viral origins of replication, host cells can be transformed with DNA controlled by appropriate expression control elements (*e.g.*, promoter, enhancer, sequences, transcription terminators, polyadenylation sites, etc.), and a selectable marker. Following the introduction of the foreign DNA, engineered cells may be allowed to grow for 1-2 days in an enriched media, and then are switched to a selective media. The selectable marker in the recombinant plasmid confers resistance to the selection and allows cells to stably integrate the plasmid into their chromosomes and grow to form foci that in turn can be cloned and expanded into cell lines. This method may advantageously be used to engineer cell lines that express the SPAN-X product. Such engineered cell lines may be particularly useful in screening and evaluation of compounds that affect the endogenous activity of the *span-x* gene product.

A number of selection systems may be used, including but not limited to the herpes simplex virus thymidine kinase (Wigler, *et al.*, 1977, Cell 11:223), hypoxanthine-guanine phosphoribosyltransferase (Szybalska and Szybalski, 1962, Proc. Natl. Acad. Sci. USA 48, 2026), and adenine phosphoribosyltransferase (Lowy, *et al.*, 1980, Cell 22, 817) genes can be employed in tk^-, $hgprt^-$ or $aprt^-$ cells, respectively. Also, antimetabolite resistance can be used as the basis of selection for the following genes: dhfr, which confers resistance to methotrexate (Wigler, *et al.*, 1980, Natl. Acad. Sci. USA 77, 3567; O'Hare, *et al.*, 1981, Proc. Natl. Acad. Sci. USA 78, 1527); gpt, which confers resistance to mycophenolic acid (Mulligan and Berg, 1981, Proc. Natl. Acad. Sci. USA 78, 2072); neo, which confers resistance to the aminoglycoside G-418 (Colberre-Garapin, *et al.*, 1981, J. Mol. Biol. 150, 1); and hygro, which confers resistance to hygromycin (Santerre, *et al.*, 1984, Gene 30, 147).

Alternatively, any fusion protein may be readily purified by utilizing an antibody specific for the fusion protein being expressed. For example, a system described by Janknecht, *et al.* allows for the ready purification of non-denatured fusion proteins

expressed in human cell lines (Janknecht, *et al.*, 1991, Proc. Natl. Acad. Sci. USA 88, 8972-8976). In this system, the gene of interest is subcloned into a vaccinia recombination plasmid such that the gene's open reading frame is translationally fused to an amino-terminal tag consisting of six histidine residues. Extracts from cells infected with recombinant vaccinia virus are loaded onto Ni^{2+}-nitriloacetic acid-agarose columns and histidine-tagged proteins are selectively eluted with imidazole-containing buffers.

The SPAN-X products can also be expressed in transgenic animals. Animals of any species, including, but not limited to, mice, rats, rabbits, guinea pigs, pigs, micro-pigs, goats, sheep, and non-human primates, *e.g.*, baboons, monkeys, and chimpanzees may be used to generate *span-x* transgenic animals. The term "transgenic," as used herein, refers to animals expressing *span-x* nucleic acid sequences from a different species (*e.g.*, mice expressing human *span-x* sequences), as well as animals that have been genetically engineered to overexpress endogenous (*i.e.*, same species) *span-x* sequences or animals that have been genetically engineered to no longer express endogenous *span-x* nucleic acid sequences (*i.e.*, "knock-out" animals), and their progeny.

The present invention also provides transgenic mice which express human wildtype *span-x* nucleic acid sequences in addition to mice engineered to express human mutant *span-x* nucleic acid sequences deleted of a functional domain, such as the nuclear translocation domain. Any technique known in the art may be used to introduce a *span-x* transgene into animals to produce the founder lines of transgenic animals. Such techniques include, but are not limited to pronuclear microinjection (Hoppe and Wagner, 1989, U.S. Pat. No. 4,873,191); retrovirus mediated gene transfer into germ lines (Van der Putten, *et al.*, 1985, Proc. Natl. Acad. Sci., USA 82, 6148-6152); gene targeting in embryonic stem cells (Thompson, *et al.*, 1989, Cell 56, 313-321); electroporation of embryos (Lo, 1983, Mol. Cell. Biol. 3, 1803-1814); and sperm-mediated gene transfer (Lavitrano *et al.*, 1989, Cell 57, 717-723; for a review of such techniques, see Gordon, 1989, Transgenic Animals, Intl. Rev. Cytol. 115, 171-229)

Any technique known in the art may be used to produce transgenic animal clones containing a *span-x* transgene, for example, nuclear transfer into enucleated oocytes of nuclei from cultured embryonic, fetal or adult cells induced to quiescence (Campbell, *et al.*, 1996, Nature 380, 64-66; Wilmut, *et al.*, Nature 385, 810-813).

The present invention provides for transgenic animals that carry a *span-x* transgene in all their cells, as well as animals that carry the transgene in some, but not all their cells, *i.e.*, mosaic animals. The transgene may be integrated as a single transgene or in concatamers, *e.g.*, head-to-head tandems or head-to-tail tandems. The transgene may also be selectively introduced into and activated in a particular cell type by following, for

example, the teaching of Lasko *et al.* (Lasko, *et al.*, 1992, Proc. Natl. Acad. Sci. USA 89, 6232-6236). The regulatory sequences required for such a cell-type specific activation will depend upon the particular cell type of interest, and will be apparent to those of skill in the art. When it is desired that the *span-x* gene transgene be integrated into the chromosomal site of the endogenous *span-x* gene, gene targeting is preferred. Briefly, when such a technique is to be utilized, vectors containing some nucleotide sequences homologous to the endogenous *span-x* gene are designed for the purpose of integrating, via homologous recombination with chromosomal sequences, into and disrupting the function of the nucleotide sequence of the endogenous *span-x* gene. The transgene may also be selectively introduced into a particular cell type, thus inactivating the endogenous *span-x* gene in only that cell type, by following, for example, the teaching of Gu, *et al.* (Gu, *et al.*, 1994, Science 265, 103-106). The regulatory sequences required for such a cell-type specific inactivation will depend upon the particular cell type of interest, and will be apparent to those of skill in the art.

Once transgenic animals have been generated, the expression of the recombinant *span-x* gene may be assayed utilizing standard techniques. Initial screening may be accomplished by Southern blot analysis or PCR techniques to analyze animal tissues to assay whether integration of the transgene has taken place. The level of mRNA expression of the transgene in the tissues of the transgenic animals may also be assessed using techniques that include, but are not limited to, Northern blot analysis of tissue samples obtained from the animal, *in situ* hybridization analysis, and RT-PCR (reverse transcriptase PCR). Samples of *span-x* gene-expressing tissue, may also be evaluated immunocytochemically using antibodies specific for the *span-x* transgene product.

5.2.2. COMPOSITIONS CONTAINING ONE OR MORE DOMAINS OF SPAN-X

In a specific embodiment, the invention provides SPAN-X fragments or analogs and derivatives of such fragments, that comprise, or alternatively consist of, one or more domains of a SPAN-X protein, for example, a nuclear translocation domain.

A specific embodiment provides molecules comprising specific fragments of SPAN-X that are those fragments in the respective SPAN-X protein most homologous to specific fragments of a human SPAN-X protein. A fragment comprising a domain of a SPAN-X homolog can be identified by protein analysis methods as described in Sections 5.3.2 or 6.

In a specific embodiment, the invention provides a fragment, derivative or analog of a SPAN-X protein that has a functional nuclear translocation domain. See above Section

5.1 for nucleic acid sequences encoding the SPAN-X nuclear localization signal. In another specific embodiment, the invention provides fusion proteins comprising a SPAN-X nuclear translocation domain that has been operatively linked to a heterologous protein. The fusion proteins are useful to cause the heterologous protein to be translocated to the nucleus. It is known that a proline followed within the next six residues by three lysines, has been shown to direct translocation of proteins into the nucleus (Hicks, G. R., and Raikhel, N. V. (1995). Protein import into the nucleus: an integrated view. *Annu. Rev. Cell. Dev. Biol.* 11, 155-158). See *infra* Section 5.6 for a more detailed discussion.

In another specific embodiment, a molecule is provided that comprises one or more domains (or functional portion thereof) of a SPAN-X protein but that also lacks one or more domains (or functional portion thereof) of a SPAN-X protein. In particular examples, SPAN-X protein derivatives are provided that lack a nuclear translocation domain.

5.3. SPAN-X ANTIBODIES

According to the invention, SPAN-X, its fragments or other derivatives, or analogs thereof, may be used as an immunogen to generate antibodies which immunospecifically bind such an immunogen. Such antibodies include, but are not limited to, polyclonal, monoclonal, chimeric, single chain, Fab fragments, and an Fab expression library. In a specific embodiment, antibodies to a human SPAN-X protein are produced. In another embodiment, antibodies to a domain (*e.g.*, the nuclear translocation domain) of a SPAN-X are produced.

Various procedures known in the art may be used for the production of polyclonal antibodies to a SPAN-X or derivative or analog. In a particular embodiment, rabbit polyclonal antibodies to an epitope of an SPAN-X encoded by a sequence of SEQ ID NOs:1 or 3 or a subsequence thereof, can be obtained. For the production of antibody, various host animals can be immunized by injection with the native SPAN-X, or a synthetic version, or derivative (*e.g.*, fragment) thereof, including but not limited to rabbits, mice, rats, etc. Various adjuvants may be used to increase the immunological response, depending on the host species, and including but not limited to Freund's (complete and incomplete), mineral gels such as aluminum hydroxide, surface active substances such as lysolecithin, pluronic polyols, polyanions, peptides, oil emulsions, keyhole limpet hemocyanins, dinitrophenol, and potentially useful human adjuvants such as BCG (bacille Calmette-Guerin) and corynebacterium parvum. For preparation of monoclonal antibodies directed toward an SPAN-X sequence or analog thereof, any technique which provides for the production of antibody molecules by continuous cell lines in culture may be used. For example, the hybridoma technique originally developed by Kohler and

Milstein (1975, Nature 256:495-497), as well as the trioma technique, the human B-cell hybridoma technique (Kozbor et al., 1983, Immunology Today 4:72), and the EBV-hybridoma technique to produce human monoclonal antibodies (Cole et al., 1985, in Monoclonal Antibodies and Cancer Therapy, Alan R. Liss, Inc., pp. 77-96). In an additional embodiment of the invention, monoclonal antibodies can be produced in germ-free animals (see PCT International Publication No. WO 89/12690, published December 12, 1989). According to the invention, human antibodies may be used and can be obtained by using human hybridomas (Cote et al., 1983, Proc. Natl. Acad. Sci. U.S.A. 80:2026-2030) or by transforming human B cells with EBV virus *in vitro* (Cole et al., 1985, in Monoclonal Antibodies and Cancer Therapy, Alan R. Liss, pp. 77-96). In fact, according to the invention, techniques developed for the production of "chimeric antibodies" (Morrison et al., 1984, Proc. Natl. Acad. Sci. U.S.A. 81:6851-6855; Neuberger et al., 1984, Nature 312:604-608; Takeda et al., 1985, Nature 314:452-454) by splicing the genes from a mouse antibody molecule specific for SPAN-X together with genes from a human antibody molecule of appropriate biological activity can be used; such antibodies are within the scope of this invention.

According to the invention, techniques described for the production of single chain antibodies (U.S. 4,946,778) can be adapted to produce SPAN-X-specific single chain antibodies. An additional embodiment of the invention utilizes the techniques described for the construction of Fab expression libraries (Huse et al., 1989, Science 246:1275-1281) to allow rapid and easy identification of monoclonal Fab fragments with the desired specificity for SPAN-Xs, derivatives, or analogs.

Antibody fragments which contain the idiotype of the molecule can be generated by known techniques. For example, such fragments include but are not limited to: the $F(ab')_2$ fragment which can be produced by pepsin digestion of the antibody molecule; the Fab' fragments which can be generated by reducing the disulfide bridges of the $F(ab')_2$ fragment, the Fab fragments which can be generated by treating the antibody molecule with papain and a reducing agent, and Fv fragments.

In the production of antibodies, screening for the desired antibody can be accomplished by techniques known in the art, *e.g.* ELISA (enzyme-linked immunosorbent assay). For example, to select antibodies which recognize a specific domain of an SPAN-X, one may assay generated hybridomas for a product which binds to an SPAN-X fragment containing such domain. For selection of an antibody that specifically binds a first SPAN-X homolog but which does not specifically bind a different SPAN-X homolog, one can select on the basis of positive binding to the first SPAN-X homolog and a lack of binding to the second SPAN-X homolog.

Antibodies specific to a domain of an SPAN-X are also provided, such as a nuclear localization domain.

The foregoing antibodies can be used in methods known in the art relating to the identification of sperm containing SPAN-X protein, separation of sperm, and the localization and activity of the SPAN-X polypeptides of the invention, *e.g.*, for imaging these proteins, in diagnostic methods, measuring levels thereof in appropriate physiological samples etc.

5.4. SPAN-X LIGANDS

Any of a variety of exogenous compounds, both naturally occurring and/or synthetic (*e.g.*, libraries of small molecules or peptides, may be screened for SPAN-X binding capacity.

All of these methods comprise the step of mixing an SPAN-X protein or fragment with test compounds, allowing time for any binding to occur, and assaying for any bound complexes. All such methods are enabled by the present disclosure of substantially pure SPAN-X proteins, substantially pure functional domain fragments, fusion proteins, antibodies, and methods of making and using the same.

In one embodiment, peptide libraries may be used to screen for agonists or antagonists of the SPAN-X of the present invention diversity libraries, such as random or combinatorial peptide or nonpeptide libraries can be screened for molecules that specifically bind to SPAN-X. Many libraries are known in the art that can be used, *e.g.*, chemically synthesized libraries, recombinant (*e.g.*, phage display libraries), and *in vitro* translation-based libraries.

Examples of chemically synthesized libraries are described in Fodor et al., 1991, Science 251:767-773; Houghten et al., 1991, Nature 354:84-86; Lam et al., 1991, Nature 354:82-84; Medynski, 1994, Bio/Technology 12:709-710; Gallop et al., 1994, J. Medicinal Chemistry 37(9):1233-1251; Ohlmeyer et al., 1993, Proc. Natl. Acad. Sci. USA 90:10922-10926; Erb et al., 1994, Proc. Natl. Acad. Sci. USA 91:11422-11426; Houghten et al., 1992, Biotechniques 13:412; Jayawickreme et al., 1994, Proc. Natl. Acad. Sci. USA 91:1614-1618; Salmon et al., 1993, Proc. Natl. Acad. Sci. USA 90:11708-11712; PCT Publication No. WO 93/20242; and Brenner and Lerner, 1992, Proc. Natl. Acad. Sci. USA 89:5381-5383.

Examples of phage display libraries are described in Scott & Smith, 1990, Science 249:386-390; Devlin et al., 1990, Science, 249:404-406; Christian, et al., 1992, J. Mol. Biol. 227:711-718; Lenstra, 1992, J. Immunol. Meth. 152:149-157; Kay et al., 1993, Gene 128:59-65; and PCT Publication No. WO 94/18318 dated August 18, 1994.

In vitro translation-based libraries include but are not limited to those described in PCT Publication No. WO 91/05058 dated April 18, 1991; and Mattheakis et al., 1994, Proc. Natl. Acad. Sci. USA 91:9022-9026.

By way of examples of nonpeptide libraries, a benzodiazepine library (*see e.g.*, Bunin et al., 1994, Proc. Natl. Acad. Sci. USA 91:4708-4712) can be adapted for use. Peptoid libraries (Simon et al., 1992, Proc. Natl. Acad. Sci. USA 89:9367-9371) can also be used. Another example of a library that can be used, in which the amide functionalities in peptides have been permethylated to generate a chemically transformed combinatorial library, is described by Ostresh et al. (1994, Proc. Natl. Acad. Sci. USA 91:11138-11142).

Screening the libraries can be accomplished by any of a variety of commonly known methods. See, *e.g.*, the following references, which disclose screening of peptide libraries: Parmley & Smith, 1989, Adv. Exp. Med. Biol. 251:215-218; Scott & Smith, 1990, Science 249:386-390; Fowlkes et al., 1992; BioTechniques 13:422-427; Oldenburg et al., 1992, Proc. Natl. Acad. Sci. USA 89:5393-5397; Yu et al., 1994, Cell 76:933-945; Staudt et al., 1988, Science 241:577-580; Bock et al., 1992, Nature 355:564-566; Tuerk et al., 1992, Proc. Natl. Acad. Sci. USA 89:6988-6992; Ellington et al., 1992, Nature 355:850-852; US 5,096,815, US 5,223,409, and US 5,198,346, all to Ladner et al.; Rebar & Pabo, 1993, Science 263:671-673; and PCT Publication No. WO 94/18318.

In a specific embodiment, screening can be carried out by contacting the library members with a SPAN-X protein (or nucleic acid or derivative) immobilized on a solid phase and harvesting those library members that bind to the protein (or nucleic acid or derivative). Examples of such screening methods, termed "panning" techniques are described by way of example in Parmley & Smith, 1988, Gene 73:305-318; Fowlkes et al., 1992, BioTechniques 13:422-427; PCT Publication No. WO 94/18318; and in references cited hereinabove.

In another embodiment, the two-hybrid system for selecting interacting proteins or peptides in yeast (Fields & Song, 1989, Nature 340:245-246; Chien et al., 1991, Proc. Natl. Acad. Sci. USA 88:9578-9582) can be used to identify molecules that specifically bind to an SPAN-X protein or derivative.

Binding interactions between two or more components can be measured in a variety of ways. One approach is to label one of the components with an easily detectable label, place it together with the other component(s) in conditions under which they would normally interact, perform a separation step which separates bound labeled component from unbound labeled component, and then measure the amount of bound component. The effect of a test agent included in the binding reaction can be determined by comparing the

amount of labeled component which binds in the presence of this agent to the amount which binds in its absence.

The separation step in this type of procedure can be accomplished in various ways. In one approach, (one of) the binding partner(s) for the labeled component can be immobilized on a solid phase prior to the binding reaction, and unbound labeled component can be removed after the binding reaction by washing the solid phase. Attachment of the binding partner to the solid phase can be accomplished in various ways known to those skilled in the art, including, but not limited to, chemical cross-linking, non-specific adhesion to a plastic surface, interaction with an antibody attached to the solid phase, interaction between a ligand attached to the binding partner (such as biotin) and a ligand-binding protein (such as avidin or streptavidin) attached to the solid phase, and so on.

Alternatively, the separation step can be accomplished after the labeled component had been allowed to interact with its binding partner(s) in solution. If the size differences between the labeled component and its binding partner(s) permit such a separation, the separation can be achieved by passing the products of the binding reaction through an ultrafilter whose pores allow passage of unbound labeled component but not of its binding partner(s) or of labeled component bound to its partner(s). Separation can also be achieved using any reagent capable of capturing a binding partner of the labeled component from solution, such as an antibody against the binding partner, a ligand-binding protein which can interact with a ligand previously attached to the binding partner, and so on.

5.5. METHODS AND COMPOSITIONS FOR DIAGNOSTIC AND PREPARATIVE USE OF SPAN-X NUCLEIC ACIDS, PROTEINS, DERIVATIVES, AND ANTIBODIES

5.5.1. METHODS AND COMPOSITIONS FOR SEPARATION OF SPAN-X-CONTAINING HUMAN SPERM AND SPERM NUCLEI

SPAN-X polynucleotides, polypeptides and antibodies may be used for SPAN-X-based purification of X-bearing sperm and Y-bearing sperm. Thus, the present invention provides methods for enriching and isolating cells that express SPAN-X protein from semen and sperm cells.

In one embodiment, the purification method is carried out by affinity chromatography, methods for which are well known in the art (*See, e.g.,* Current Protocols in Molecular Biology, Vol. 2, Ed. Ausubel et al., Greene Publish. Assoc. & Wiley Interscience). A solid phase surface comprising a specific partner molecule is used to specifically capture and remove the SPAN-X-bearing components from a mixture of cells or nuclei. Such a partner molecule may comprise a primary or secondary antibody, a

ligand, or an affinity tag attached to a ligand or an antibody. In various embodiments, the solid phase surface may be, but is not restricted to, polycarbonate, polystyrene, polypropylene, polyethlene, glass, nitrocellulose, dextran, nylon, polyacrylamide and agarose. The support configuration can include beads, membranes, microparticles, the interior surface of a reaction vessel such as a microtiter plate, test tube or other reaction vessel.

In a preferred embodiment, an affinity column is composed of anti-SPAN-X antibody or antiserum stabilized on an appropriate matrix, such as a sepharose column. X- and Y- bearing sperm cells or nuclei are contacted with the column, under conditions that allow immunospecific binding to occur. The material that does not bind to the column is collected, and can be re-loaded on the column. This process may be repeated several times, as desired. The affinity column is then washed, and the material retained on the column is eluted. The material that bound to the column, containing X-bearing sperm or nuclei and Y-bearing sperm or nuclei, is eluted and collected in separate containers, to be used in the fertilization techniques, described hereinbelow.

In various other embodiments, to facilitate recovery or purification of SPAN-X cells or nuclei, the partner molecule, *i.e.*, an antibody or a ligand, may be labeled. In one embodiment, antibodies against SPAN-X proteins may be labeled by conjugation of an affinity compound to such antibodies. Affinity compounds that can be used include but are not limited to biotin, photobiotin, fluorescein isothiocyante (FITC), or phycoerythrin (PE), or other compounds known in the art. Cells or nuclei retaining labeled antibodies are then separated from cells that do not bind such antibodies by techniques known in the art. In one embodiment, affinity compounds or affinity tags can be conjugated to the antibodies through a polyfunctional crosslinker, and preferably a bifunctional molecule. As used herein the term polyfunctional crosslinker encompasses molecules having more than one functional group that reacts with a functional group on the antibody. Typically, such crosslinker forms covalent bonds with an amino or sulfhydryl group on a polypeptide. For example, biotin N-hydroxysuccinimide esters may be used.

In another embodiment, the label may be a peptide tag, i.e., a fusion protein, may be used to attach an antibody or ligand to a solid phase support. Such a fusion protein can be made by ligating a ligand gene sequence to the sequence encoding the peptide tag in the proper reading frame. A variety of peptide tags known in the art may be used, such as but not limited to the immunoglobulin constant regions, polyhistidine sequence (Petty, 1996, Metal-chelate affinity chromatography (Current Protocols in Molecular Biology, Vol. 2, Ed. Ausubel et al., Greene Publish. Assoc. & Wiley Interscience), glutathione S-transferase (GST; Smith, 1993, Methods Mol. Cell Bio. 4:220-229), the *E. coli* maltose binding protein

(Guan et al., 1987, Gene 67:21-30), and various cellulose binding domains (US 5,496,934; US 5,202,247; US 5,137,819; Tomme et al., 1994, Protein Eng. 7:117-123), etc. Other peptide tags may impart fluorescent properties to a ligand or antibody, e.g., portions of green fluorescent protein and the like. Other possible peptide tags are short amino acid sequences to which monoclonal antibodies are available, such as but not limited to the following well known examples, the FLAG epitope, the myc epitope at amino acids 408-439, the influenza virus hemagglutinin (HA) epitope. Other peptide tags are recognized by specific binding partners and thus facilitate isolation by affinity binding to the binding partner which can be immobilized onto a solid support.

In another embodiment, sperm cells or nuclei may be sorted using a using a fluorescence activated cell sorter (FACS). Fluorescence activated cell sorting (FACS) is a well-known method for separating particles, including cells, based on the fluorescent properties of the particles (Kamarch, 1987, Methods Enzymol, 151:150-165). Laser excitation of fluorescent moieties in the individual particles results in a small electrical charge allowing electromagnetic separation of positive and negative particles from a mixture. In one embodiment, SPAN-X antibody is labeled with a fluorescent label. A sample of sperm cells or sperm nuclei is incubated with the fluorescently-labeled antibodies or ligands for a time period sufficient to allow the labeled antibodies or ligands to bind to cells, between 10 and 60 minutes. Cells or nuclei are processed through a cell sorter, thereby purifying SPAN-X containing cells. Cells or nuclei are eluted from the beads, allowing recovery or purification of SPAN-X-containing cells. FACS sorted particles may be directly deposited into a recovery vesicle, such as individual wells of 96-well or 384-well plates to facilitate recovery or purification.

In another embodiment, magnetic beads can be used to separate cells or nuclei. SPAN-X- containing cells may be purified using a using a magnetic activated cell sorting (MACS) technique, a method for separating particles based on their ability to bind magnetic beads (0.5-100µm diameter) (Dynal, 1995). A variety of useful modifications can be performed on the magnetic microspheres, including covalent addition of antibody which specifically recognizes SPAN-X. A magnetic field is then applied, to physically manipulate the selected beads. In a specific embodiment, antibodies to SPAN-X are coupled to magnetic beads. The beads are then mixed with the sperm cells or nuclei to allow binding. Cells are then passed through a magnetic field to isolate X-bearing and Y-bearing sperm cells or nuclei.

The efficiency of recovery or purification procedures can be monitored using SPAN-X polynucleotides, polypeptides, antibodies or kits provided by the invention and described in Section 5.4.1, hereinabove.

Kits for purification purposes are also provided. Such kits comprise in one or more containers an anti-SPAN-X antibody, and a suitable affinity matrix. Optionally, a labeled binding partner to the antibody is also provided. A kit can optionally further comprise in a container a predetermined amount of a purified SPAN-X protein or nucleic acid, *e.g.*, for use as a standard or control, and a pair of primers (*e.g.*, each in the size range of 6-30 nucleotides) that are capable of priming amplification under appropriate reaction conditions of at least a portion of a *span-x* nucleic acid, by any of the variety of methods known in the art.

5.5.2. METHODS FOR THE USE OF PURIFIED SPAN-X-CONTAINING SPERM AND SPERM NUCLEI FOR FERTILIZATION

SPAN-X-containing X- and Y- bearing sperm, purified using the methods described hereinabove, may be used for fertilization using artificial insemination, intrauterine insemination or *in-vitro* fertilization techniques that are known in the art. In an embodiment, the purified sperm are directly injected into the egg, by a technique known as intracytoplasmic sperm injection (ICSI) or subzonal insemination (SUZI) (Van Steirteghem *et al.*, 1994, Reprod. Fertil. Dev. 6: 85-91; Van Steirteghem *et al.*, 1994, Baillieres Clin. Obstet. Gynaecol. 8:85-93). This method is preferable when using sperm purified by anti-SPAN-X affinity chromatography, since the efficiency of fertilization may be greatly improved over other methods known in the art. Furthermore, using this technique, viable sperm may not be required to obtain fertilized ovum.

5.5.3. METHODS AND COMPOSITIONS FOR THE USE OF SPAN-X AS A MARKER TO SEPARATE FERTILIZATION COMPETENT AND INCOMPETENT SPERMATOZOA

SPAN-X proteins, analogues, derivatives, and subsequences thereof, *span-x* nucleic acids (and sequences complementary thereto), anti-SPAN-X antibodies, have uses in diagnostics. The SPAN-X and *span-x* nucleic acids can be used in diagnostic assays as markers for fertilization competency of spermatozoa. Such assays are useful, for example, for identifying the presence of assessing fertilization competency of spermatozoa or for monitoring SPAN-X in developing spermatazoa.

The molecules of the present invention can be used in assays, such as immunoassays, to detect, prognose, diagnose, or monitor sperm cells expressing SPAN-X. In particular, such an immunoassay is carried out by a method comprising contacting a sample derived from a subject with an anti-SPAN-X antibody under conditions such that immunospecific binding can occur, and detecting or measuring the amount of any

immunospecific binding by the antibody. In another specific aspect, such binding of antibody, in tissue sections, can be used to detect aberrant SPAN-X localization or aberrant (*e.g.*, low or absent) levels of SPAN-X in a subject having a particular disorder, such as infertility. In a specific embodiment, antibody to SPAN-X can be used to assay a patient tissue or serum sample for the presence of SPAN-X where an aberrant level of SPAN-X is an indication of a diseased condition. By "aberrant levels," is meant increased or decreased levels relative to that present, or a standard level representing that present, in a sample of a subject not having the disorder.

The immunoassays which can be used include, but are not limited to, competitive and non-competitive assay systems using techniques such as Western blots, immunohistochemistry radioimmunoassays, ELISA (enzyme linked immunosorbent assay), "sandwich" immunoassays, immunoprecipitation assays, precipitin reactions, gel diffusion precipitin reactions, immunodiffusion assays, agglutination assays, complement-fixation assays, immunoradiometric assays, fluorescent immunoassays, protein A immunoassays, to name but a few (for immunoassay techniques, see, for example, Harlow and Lane, eds., 1988, "Antibodies: A Laboratory Manual", Cold Spring Harbor Press, Cold Spring Harbor).

Span-x genes and related nucleic acid sequences and subsequences, including complementary sequences, can also be used in hybridization assays. *Span-x* nucleic acid sequences, or subsequences thereof comprising about at least 8 nucleotides, can be used as hybridization probes. Hybridization assays can be used to detect, prognose, diagnose, or monitor conditions, disorders, or disease states associated with aberrant changes in *span-x* expression and/or activity as described *supra*. In particular, such a hybridization assay is carried out by a method comprising contacting a sample containing nucleic acid with a nucleic acid probe capable of hybridizing to *span-x* DNA or RNA, under conditions such that hybridization can occur, and detecting or measuring any resulting hybridization.

In specific embodiments, the presence of the *span-x* gene product can be diagnosed, or the suspected presence can be screened for, or a predisposition to develop such disorders can be detected, by detecting decreased or increased levels of SPAN-X protein, *span-x* RNA, or SPAN-X functional activity (*e.g.*, binding to SPAN-X specific antibodies). Abnormalities of such SPAN-X proteins may also be diagnosed by detecting mutations in *span-x* RNA, DNA or SPAN-X protein (*e.g.*, translocations in SPAN-X nucleic acids, truncations in the SPAN-X gene or protein, changes in nucleotide or amino acid sequence relative to wild-type SPAN-X) that cause decreased expression or activity of SPAN-X. By way of example, levels of SPAN-X protein can be detected by immunoassay, levels of *span-x* RNA can be detected by hybridization assays (*e.g.*, Northern blots, *in situ*-hybridization), translocations, deletions and point mutations in SPAN-X nucleic acids can

be detected by Southern blotting, FISH, RFLP analysis, SSCP, PCR using primers that preferably generate a fragment spanning at least most of the SPAN-X gene, sequencing of *span-x* genomic DNA or cDNA obtained from the patient, etc.

In a preferred embodiment, levels of *span-x* mRNA or protein in a sample from a subject are detected or measured, in which decreased levels indicate that the subject has, or has a predisposition to developing, for example an X-linked fertility disorder.

Kits for diagnostic use are also provided, that comprise in one or more containers an anti-SPAN-X antibody, and, optionally, a labeled binding partner to the antibody. Alternatively, the anti-SPAN-X antibody can be labeled (with a detectable marker, *e.g.*, a chemiluminescent, enzymatic, fluorescent, or radioactive moiety). A kit is also provided that comprises in one or more containers a nucleic acid probe capable of hybridizing to *span-x* RNA. In a specific embodiment, a kit can comprise in one or more containers a pair of primers (*e.g.*, each in the size range of 6-30 nucleotides) that are capable of priming amplification [*e.g.*, by polymerase chain reaction (see *e.g.*, Innis et al., 1990, PCR Protocols, Academic Press, Inc., San Diego, CA), ligase chain reaction (see EP 320,308) use of Qβ replicase, cyclic probe reaction, or other methods known in the art] under appropriate reaction conditions of at least a portion of a *span-x* nucleic acid. A kit can optionally further comprise in a container a predetermined amount of a purified SPAN-X protein or nucleic acid, *e.g.*, for use as a standard or control.

5.6. METHODS AND COMPOSITIONS FOR THERAPEUTIC USE OF SPAN-X NUCLEIC ACIDS, POLYPEPTIDES, DERIVATIVES, AND MODULATORS

5.6.1. THERAPEUTIC USE OF SPAN-X REGULATORY SEQUENCES FOR TARGETING GENES TO THE TESTIS

As exemplified in Section 6, analysis of human/mouse somatic cell hybrids containing various complements of human chromosomes has allowed the assignment of the human *span-x* gene to chromosome Xq22-q28. FISH analysis of male metaphase chromosomes is consistent with this assignment and further refines the span-*x* locus to chromosome Xq27.1. The homology between the SPAN-X cDNA and the human PAC clone mapped to Xq26.3-Xq27.1 confirms these results and has allowed mapping of SPAN-X gene and promoter nucleotide sequences. This Southern analysis indicates that SPAN-X is a single copy gene, and Northern analysis of *span-x* gene transcripts indicate that *span-x* is specifically transcribed in the testis, and probably regulated at the transcriptional level.

Span-x promoter sequences can be used advantageously to drive spermatid-specific expression of heterologous gene products. A vector comprising the *span-x* promoter nucleotide sequences operably linked to a heterologous gene can be useful for gene therapy and contraception. In one embodiment, these sequences can be used for contraceptive or

sterilization purposes. SPAN-X promoter sequences can be inserted into a vector operatively linked to a gene that will kill the cell in which it is expressed. Examples of such genes are known in the art, including, but not limited to, spermicides and toxins. In another embodiment, such a vector can be used to target cancer cells of a patient with a testes-specific proliferative disorder or cancer to inhibit growth or kill the cancer cell.

In one embodiment, SPAN-X promoter sequences can be used to drive spermatid-specific expression of drugs or toxins using gene therapy techniques in cells of a patient with a testes-specific proliferative disorder or cancer to inhibit growth or kill the cancer cell. In another embodiment, gene therapy techniques using promoter constructs to drive spermatid-specific expression of drugs or toxins can be used for sterilization or contraception in the testis.

The genomic sequence of the *span-x* gene contains regulatory sequences both in the non-coding 5′-flanking gene of polynucleotide sequence of SEQ ID NO:21 or SEQ ID NO:22 can be assessed by any known method. Methods for *span-x* gene comprise the polynucleotide sequences located between the nucleotide in position -2000 and the nucleotide in position +103 of the nucleotide sequence of SEQ ID NO:21 or SEQ ID NO:22 or more preferably between positions -2074 and +103 of SEQ ID NO:21 or SEQ ID NO:22.

The promoter activity of the regulatory regions contained in the region and in the non-coding 3′-flanking region that border the *span-x* transcribed region containing the 2 exons of this gene. 5′-regulatory sequences of the *span-x* transcribed region containing the 2 exons of this gene. 5′-regulatory sequences of the *span-x* identifying the polynucleotide fragments of SEQ ID NO:21 or SEQ ID NO:22 involved in the regulation of the expression of the *span-x* gene are well-known to those skilled in the art (see Sambrook et al., Molecular Cloning A Laboratory Manual, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY, 1989). An example of a typical method that can be used involves a recombinant vector carrying a reporter gene and genomic sequences from the *span-x* genomic promoter sequence of SEQ ID NO. 21 or SEQ ID NO. 22. Briefly, the expression of the reporter gene (for example, green fluorescent protein, luciferase, β-galactosidase, or chloramphenicol acetyl transferase) is detected when placed under the control of a biologically active polynucleotide fragment. Genomic sequences located upstream of the first exon of the gene may be cloned into any suitable promoter reporter vector, such as the pSEAPBasic, pSEAP-Enhancer, pβgal-Basic, pβgal-Enhancer, or pEGFP-1 Promoter Reporter vectors available from Clontech, or pGL2-basic or pGL3-basic promoterless luciferase reporter gene vector from Promega. Each of these promoter reporter vectors include multiple cloning sites positioned upstream of a reporter gene encoding a readily assayable protein such as secreted alkaline phosphatase, green fluorescent protein,

luciferase, or β-galactosidase. The sequences upstream of the first *span-x* exon are inserted into the cloning sites upstream of the reporter gene in both orientations and introduced into an appropriate host cell. The level of reporter protein is assayed and compared to the level obtained with a vector lacking an insert in the cloning site. The presence of an elevated expression level in the vector containing the insert with respect to the control vector indicates the presence of a promoter in the insert.

Promoter sequences within the 5' non-coding regions of the *span-x* gene may be further defined by constructing nested 5' and/or 3' deletions using conventional techniques such as Exonuclease III or appropriate restriction endonuclease digestion. The resulting deletion fragments can be inserted into the promoter reporter vector to determine whether the deletion has reduced or obliterated promoter activity, such as described, for example, by Coles et al. (Hum. Mol. Genet., 7:791-800, 1998). In this way, the boundaries of the promoters may be defined. If desired, potential individual regulatory sites within the promoter may be identified using site directed mutagenesis or linker scanning to obliterate potential transcription factor binding sites within the promoter individually or in combination. The effects of these mutations on transcription levels may be determined by inserting the mutations into cloning sites in promoter reporter vectors. This type of assays are well known to those skilled in the art (WO 97/17359, US 5,374,544, EP 582 796, US 5,698,389, US 5,643,746, US5,502,176, and US 5,266,488).

The activity and the specificity of the promoter of the *span-x* gene can further be assessed by monitoring the expression level of a detectable polynucleotide operably linked to the SPAN-X promoter in different types of cells and tissues. The detectable polynucleotide may be either a polynucleotide that specifically hybridizes with a predefined oligonucleotide probe, or a polynucleotide encoding a detectable protein, including a *span-x* polypeptide or a fragment or a variant thereof. This type of assay is well known to those skilled in the art (US 5,502,176 and US 5,266,488).

Polynucleotides carrying the regulatory elements located both at the 5' end and at the 3' end of the *span-x* gene coding region may be advantageously used to control the transcriptional and translational activity of an heterologous polynucleotide of interest, said polynucleotide being heterologous as regards to the *span-x* regulatory region.

Thus, the present invention also provides a purified, isolated, and recombinant nucleic acid comprising a polynucleotide sequence located between the nucleotide in position -2000 and the nucleotide in position +103 of the nucleotide sequence of SEQ ID NO. 21 or SEQ ID NO 22, or more preferably between positions -2074 and +103 of SEQ ID NO. 21 or SEQ ID NO: 22; or a sequence complementary thereto or a functionally active fragment thereof.

By a "functionally active" fragment of SEQ ID NO. 21 or SEQ ID NO. 22 according to the present invention is intended a polynucleotide comprising or alternatively consisting of a fragment of said polynucleotide which is functional as a regulatory region for expressing a recombinant polypeptide or a recombinant polynucleotide in a recombinant cell host. For the purpose of the invention, a nucleic acid or polynucleotide is "functional" as a regulatory region for expressing a recombinant polypeptide or a recombinant polynucleotide if said regulatory polynucleotide contains nucleotide sequences which contain transcriptional and translational regulatory information, and such sequences are "operably linked" to nucleotide sequences which encode the desired polypeptide or the desired polynucleotide.

The regulatory polynucleotides according to the invention may be advantageously part of a recombinant expression vector that may be used to express a coding sequence in a desired host cell or host organism.

5.6.2. GENE REPLACEMENT DELIVERY OR THERAPY

With respect to testes-specific gene expression, *span-x* gene regulatory sequences, described, above, in Section 5.6.1 can, for example, be utilized for the treatment of proliferative disorders such as testicular cancer. Such treatment can be administered, for example, in the form of gene replacement therapy. Specifically, one or more copies of a normal *span-x* gene or a portion of the *span-x* gene that directs the production of a heterologous gene product that is toxic to the cell, may be inserted into the appropriate cells within a patient, using vectors that include, but are not limited to, adenovirus, adeno-associated virus, and retrovirus vectors, in addition to other particles that introduce DNA into cells, such as liposomes. Such gene replacement therapy techniques should be capable delivering *span-x* gene sequences to testis-specific cell types within patients.

In another embodiment, techniques for delivery involve direct administration of such *span-x* gene sequences to the site of the cells in which the *span-x* nucleic acid sequences are to be expressed.

Alternatively, cells, preferably autologous cells, can be engineered to express *span-x* nucleic acid sequences, and may then be introduced into a patient in positions appropriate for the amelioration of a testes-specific disorder, such as proliferative or differentiative disorders, *e.g.*, cancer and tumorigenesis. The expression of the heterologous gene sequences is controlled by the appropriate *span-x* gene regulatory sequences to allow such expression in testes cells. When the cells to be administered are non-autologous cells, they can be administered using well known techniques that prevent a host immune response against the introduced cells from developing. For example, the cells may be introduced in

an encapsulated form which, while allowing for an exchange of components with the immediate extracellular environment, does not allow the introduced cells to be recognized by the host immune system.

As described in more detail in Section 6 below, motif analysis identified two overlapping consensus nuclear localization signals at amino acids 37-43 (PAPKKMK)(SEQ ID NO:15) and 39-45 (PKKMKTS)(SEQ ID NO: 16) of the SPAN-Xa ORF (FIG. 2C). These amino acids are conserved in the SPAN-Xb peptide sequence. The overlapping or embedded nuclear localization signal sequence of the SPAN-Xa ORF (FIG. 2C) has the amino acid sequence spanning amino acids 37-45 (PAPKKMKTS)(SEQ ID NO:17). The embedded consensus nuclear localization signal sequence is also conserved in the SPAN-Xb peptide sequence.

As indicated above in Sections 5.1 and 5.2, the present invention encompasses the nucleic acid sequences encoding the *span-x* nuclear localization segnal sequences, SPAN-X nuclear localization signal sequence polypeptides, and methods for using the nuclear localization signal sequences for gene delivery or therapy.

Thus, the invention also encompasses isolated nucleic acid molecules comprising the DNA sequence of the consensus nuclear localization signal described above in operative association with a nucleic acid encoding heterologous polynucleotide. The invention also encompasses isolated fusion proteins comprising the SPAN-X consensus nuclear localization signals operatively associated with a heterologous polypeptide.

The invention also encompasses a method of importing a biologically active molecule into a cell *in vitro*, *in vivo,* or *ex vivo* comprising administering to the cell a complex comprising the molecule linked to an importation competent nuclear localization signal sequence, thereby importing the molecule into the nucleus of the cell. In particular, the invention encompasses a method of importing a biologically active molecule into a cell *in vitro*, *in vivo,* or *ex vivo* comprising administering to the cell a complex comprising the molecule linked to the nuclear localization signal sequence comprising the amino acid sequence set forth in SEQ ID NO:15 (PAPKKMK), SEQ ID NO:16 (PKKMKTS), or SEQ ID NO:17 (PAPKKMKTS), and wherein the presence of any one or more of the aforementioned nuclear localization signal sequences serves to import the biologically active molecule into the nucleus of the cell.

The invention also encompasses a method of importing a biologically active molecule into a cell *in vitro*, *in vivo,* or *ex vivo* comprising administering to the cell a complex comprising the molecule linked to an importation competent nuclear localization signal sequence, wherein the presence of the importation competent nuclear localization

signal sequence results in an enhanced or more rapid transfer of the biologically active molecule into the nucleus of the cell.

With respect to testes-specific gene therapy, the SPAN-X nuclear localization signal sequences of the invention can be used, for example, in gene targeting heterologous sequences to the nucleus of spermatozoa. The invention further encompasses the use of the SPAN-X nuclear localization signal sequences described above in gene targeting heterologous sequences to the redundant nuclear membrane and the nuclear vacuoles in mature spermatozoa.

In one embodiment, the invention encompasses the use of SPAN-X promoter sequences either alone, or in combination with the nucleic acid sequences encoding the nuclear localization signal sequences set forth in SEQ ID NO:18, SEQ ID NO: 19, or SEQ ID NO:20 , respectively, which can be inserted into a vector operatively linked to a gene that will kill the cell in which it is expressed. Examples of such genes are known in the art, including, but not limited to, spermicides and toxins. In another embodiment, such a vector can be used to target cancer cells of a patient with a testes-specific proliferative disorder or cancer to inhibit growth or kill the cancer cell.

In one embodiment, the invention also encompasses the use of SPAN-X promoter sequences either alone, or in combination with any nucleic acid sequences encoding the nuclear localization signal sequences set forth in SEQ ID NO: 16 (PKKMKTS), SEQ ID NO:17 (PAPKKMKTS), or SEQ ID NO:15 (PAPKKMK), respectively, in which the SPAN-X promoter sequences can be used to drive spermatid-specific expression of drugs or toxins using gene therapy techniques in cells of a patient with a testes-specific proliferative disorder or cancer to inhibit growth or kill the cancer cell. In another embodiment, gene therapy techniques using promoter constructs either alone, or in combination with the nucleic acid sequences encoding the nuclear localization signal sequences set forth in SEQ ID NO:16 (PKKMKTS), SEQ ID NO:17 (PAPKKMKTS), or SEQ ID NO:15 (PAPKKMK), respectively, can be used to drive spermatid-specific expression of drugs or toxins can be used for sterilization or contraception in the testis.

5.6.3. TARGET DISORDERS

With respect to specific diseases and disorders for gene therapy diseases that can be treated or prevented by the methods of the present invention include, but are not limited to: diseases and disorders involving a deficiency in cell proliferation or in which cell proliferation is desired for treatment or prevention, and that can be treated or prevented by introduction of a heterologous gene in a testes-specific manner, include, but are not limited to, degenerative disorders, growth deficiencies, hypoproliferative disorders, physical

trauma, lesions, and wounds; for example, to promote wound healing, or to promote regeneration in degenerated, lesioned or injured tissues, etc. In a specific embodiment, testicular disorders are treated. Other disorders that are contemplated within the scope of the invention are fertility disorders. Gene therapy for contraceptive or sterilization of otherwise normal patients or subjects for veterinary purposes are also within the scope of the invention. The subject is preferably an animal, including, but not limited to, animals such as foxes, rabbits, rodents, cows, pigs, horses, chickens, cats, dogs, etc., and is preferably a mammal, and most preferably human. In a specific embodiment, a non-human mammal is the subject.

5.6.5. PHARMACEUTICAL PREPARATIONS AND METHODS OF ADMINISTRATION

Formulations and methods of administration that can be employed when the therapeutic comprises a nucleic acid are described in Sections 5.1-5.6 above; additional appropriate formulations and routes of administration can be selected from among those described hereinbelow.

Various delivery systems are known and can be used to administer a therapeutic of the invention, *e.g.*, encapsulation in liposomes, microparticles, microcapsules, recombinant cells capable of expressing the therapeutic, receptor-mediated endocytosis (see, *e.g.*, Wu and Wu, 1987, J. Biol. Chem. 262:4429-4432), construction of a therapeutic nucleic acid as part of a retroviral or other vector, etc. Methods of introduction include but are not limited to intradermal, intramuscular, intraperitoneal, intravenous, subcutaneous, intranasal, epidural, and oral routes. The compounds may be administered by any convenient route, for example by infusion or bolus injection, by absorption through epithelial or mucocutaneous linings (*e.g.*, oral mucosa, rectal and intestinal mucosa, etc.) and may be administered together with other biologically active agents. Administration can be systemic or local. In addition, it may be desirable to introduce the pharmaceutical compositions of the invention into sperm cells by any suitable route, including intraventricular and intrathecal injection; intraventricular injection may be facilitated by an intraventricular catheter, for example, attached to a reservoir, such as an Ommaya reservoir. Pulmonary administration can also be employed, *e.g.*, by use of an inhaler or nebulizer, and formulation with an aerosolizing agent.

In a specific embodiment, it may be desirable to administer the pharmaceutical compositions of the invention locally to the area in need of treatment; this may be achieved by, for example, and not by way of limitation, local infusion during surgery, topical application, *e.g.*, in conjunction with a wound dressing after surgery, by injection, by means

of a catheter, by means of a suppository, or by means of an implant, said implant being of a porous, non-porous, or gelatinous material, including membranes, such as sialastic membranes, or fibers. In one embodiment, administration can be by direct injection at the site (or former site) of a malignant tumor or neoplastic or pre-neoplastic tissue.

In another embodiment, the therapeutic can be delivered in a vesicle, in particular a liposome (*see* Langer, Science 249:1527-1533 (1990); Treat et al., in Liposomes in the Therapy of Infectious Disease and Cancer, Lopez-Berestein and Fidler (eds.), Liss, New York, pp. 353-365 (1989); Lopez-Berestein, *ibid.*, pp. 317-327; *see* generally *ibid.*)

In yet another embodiment, the therapeutic can be delivered in a controlled release system. In one embodiment, a pump may be used (*see* Langer, *supra*; Sefton, CRC Crit. Ref. Biomed. Eng. 14:201 (1987); Buchwald et al., Surgery 88:507 (1980); Saudek et al., N. Engl. J. Med. 321:574 (1989)). In another embodiment, polymeric materials can be used (*see* Medical Applications of Controlled Release, Langer and Wise (eds.), CRC Pres., Boca Raton, Florida (1974); Controlled Drug Bioavailability, Drug Product Design and Performance, Smolen and Ball (eds.), Wiley, New York (1984); Ranger and Peppas, J. Macromol. Sci. Rev. Macromol. Chem. 23:61 (1983); *see also* Levy et al., Science 228:190 (1985); During et al., Ann. Neurol. 25:351 (1989); Howard et al., J. Neurosurg. 71:105 (1989)). In yet another embodiment, a controlled release system can be placed in proximity of the therapeutic target, *i.e.*, the testes, thus requiring only a fraction of the systemic dose (*see, e.g.*, Goodson, 1984, in Medical Applications of Controlled Release, *supra*, vol. 2, pp. 115-138).

Other controlled release systems are discussed in the review by Langer (1990, Science 249:1527-1533).

In a specific embodiment where the therapeutic is a nucleic acid encoding a protein therapeutic, the nucleic acid can be administered *in vivo* to promote expression of its encoded protein, by constructing it as part of an appropriate nucleic acid expression vector and administering it so that it becomes intracellular, *e.g.*, by use of a retroviral vector (see US 4,980,286), or by direct injection, or by use of microparticle bombardment (*e.g.*, a gene gun; Biolistic, Dupont), or coating with lipids or cell-surface receptors or transfecting agents, or by administering it in linkage to a homeobox-like peptide which is known to enter the nucleus (see *e.g.*, Joliot et al., 1991, Proc. Natl. Acad. Sci. USA 88:1864-1868), etc. Alternatively, a nucleic acid therapeutic can be introduced intracellularly and incorporated within host cell DNA for expression, by homologous recombination.

The present invention also provides pharmaceutical compositions. Such compositions comprise a therapeutically effective amount of a therapeutic, and a pharmaceutically acceptable carrier. In a specific embodiment, the term "pharmaceutically

acceptable" means approved by a regulatory agency of the Federal or a state government or listed in the U.S. Pharmacopeia or other generally recognized pharmacopeia for use in animals, and more particularly in humans. The term "carrier" refers to a diluent, adjuvant, excipient, or vehicle with which the therapeutic is administered. Such pharmaceutical carriers can be sterile liquids, such as water and oils, including those of petroleum, animal, vegetable or synthetic origin, such as peanut oil, soybean oil, mineral oil, sesame oil and the like. Water is a preferred carrier when the pharmaceutical composition is administered intravenously. Saline solutions and aqueous dextrose and glycerol solutions can also be employed as liquid carriers, particularly for injectable solutions. Suitable pharmaceutical excipients include starch, glucose, lactose, sucrose, gelatin, malt, rice, flour, chalk, silica gel, sodium stearate, glycerol monostearate, talc, sodium chloride, dried skim milk, glycerol, propylene, glycol, water, ethanol and the like. The composition, if desired, can also contain minor amounts of wetting or emulsifying agents, or pH buffering agents. These compositions can take the form of solutions, suspensions, emulsion, tablets, pills, capsules, powders, sustained-release formulations and the like. The composition can be formulated as a suppository, with traditional binders and carriers such as triglycerides. Oral formulation can include standard carriers such as pharmaceutical grades of mannitol, lactose, starch, magnesium stearate, sodium saccharine, cellulose, magnesium carbonate, etc. Examples of suitable pharmaceutical carriers are described in "Remington's Pharmaceutical Sciences" by E.W. Martin. Such compositions will contain a therapeutically effective amount of the therapeutic, preferably in purified form, together with a suitable amount of carrier so as to provide the form for proper administration to the patient. The formulation should suit the mode of administration.

In a preferred embodiment, the composition is formulated in accordance with routine procedures as a pharmaceutical composition adapted for intravenous administration to human beings. Typically, compositions for intravenous administration are solutions in sterile isotonic aqueous buffer. Where necessary, the composition may also include a solubilizing agent and a local anesthetic such as lignocaine to ease pain at the site of the injection. Generally, the ingredients are supplied either separately or mixed together in unit dosage form, for example, as a dry lyophilized powder or water free concentrate in a hermetically sealed container such as an ampoule or sachette indicating the quantity of active agent. Where the composition is to be administered by infusion, it can be dispensed with an infusion bottle containing sterile pharmaceutical grade water or saline. Where the composition is administered by injection, an ampoule of sterile water for injection or saline can be provided so that the ingredients may be mixed prior to administration.

The therapeutics of the invention can be formulated as neutral or salt forms. Pharmaceutically acceptable salts include those formed with free amino groups such as those derived from hydrochloric, phosphoric, acetic, oxalic, tartaric acids, etc., and those formed with free carboxyl groups such as those derived from sodium, potassium, ammonium, calcium, ferric hydroxides, isopropylamine, triethylamine, 2-ethylamino ethanol, histidine, procaine, etc.

The amount of the therapeutic of the invention which will be effective in the treatment of a particular disorder or condition will depend on the nature of the disorder or condition, and can be determined by standard clinical techniques. In addition, *in vitro* assays may optionally be employed to help identify optimal dosage ranges. The precise dose to be employed in the formulation will also depend on the route of administration, and the seriousness of the disease or disorder, and should be decided according to the judgment of the practitioner and each patient's circumstances. However, suitable dosage ranges for intravenous administration are generally about 20-500 micrograms of active compound per kilogram body weight. Suitable dosage ranges for intranasal administration are generally about 0.01 pg/kg body weight to 1 mg/kg body weight. Effective doses may be extrapolated from dose-response curves derived from *in vitro* or animal model test systems.

Suppositories generally contain active ingredient in the range of 0.5% to 10% by weight; oral formulations preferably contain 10% to 95% active ingredient.

The invention also provides a pharmaceutical pack or kit comprising one or more containers filled with one or more of the ingredients of the pharmaceutical compositions of the invention. Optionally associated with such container(s) can be a notice in the form prescribed by a governmental agency regulating the manufacture, use or sale of pharmaceuticals or biological products, which notice reflects approval by the agency of manufacture, use or sale for human administration.

6. EXAMPLE: IDENTIFICATION OF X-LINKED SPERMATID-SPECIFIC EXPRESSION OF A PROTEIN ASSOCIATED WITH THE NUCLEUS (SPAN-X) AND PRESENT IN 50% OF HUMAN SPERMATOZOA

Described herein, are the methods used to identify and characterize SPAN-X, (sperm protein associated with the nucleus on the X chromosome), the first protein marker localized to human sperm nuclear vacuoles and redundant nuclear membranes. The gene encoding this novel protein maps to chromosome X and exhibits testis-specific, post-meiotic expression. Antibodies against recombinant SPAN-X recognize a cluster of approximately 19 acidic protein spots migrating between 20-23 kDa on two-dimensional immunoblots and localize SPAN-X to nuclear craters and cytoplasmic droplets by

immunofluorescence microscopy. Ultrastructurally, SPAN-X localizes to irregular membrane structures within sperm nuclear vacuoles and to the redundant nuclear membranes. Significantly, 50% of ejaculated human spermatozoa exhibit immunostaining for SPAN-X, an X-linked, haploid-expressed gene product that is equally distributed between X- and Y-bearing spermatozoa.

6.1. MATERIALS AND METHODS

Monoclonal Antibodies

The HSA-5 and HSA-6 hybridomas were produced by fusing splenocytes from BALB/c mice injected with ionophore-treated acrosome reacted human spermatozoa in the laboratory of Chi-Yu Gregory Lee (University of British Columbia, Vancouver, Canada; Lee, C. Y., Wong, E., The, C. Z., and Nishizawa, Y. (1985). Generation of mouse oocyte monoclonal antibodies: their effects and those of antisperm monoclonal antibodies on in vitro fertilization. *J. Reprod. Immunol.* 7, 3-13). The IgM class monoclonal antibodies (mAbs) HSA-5 and HSA-6 (Fichorova R., and Anderson, D. J. (1991). Use of sperm viability and acrosomal status assays in combination with immunofluorescence technique to ascertain surface expression of sperm antigens. *J. Reprod. Immunol.* 20, 1-20) were designated S71 and 572, respectively, by the World Health Organization workshop on anti-sperm antibodies (Anderson, D. J. (1990). World Health Organization sperm antigen workshops and sperm antigen project. In "Gamete Interaction: Prospects for Immunocontraception" (N. J. Alexander, D. Griffin, J. M. Spieler, and G. M. H. Waites, Eds.), pp. 103-109. Wiley-Liss, Inc., New York). The HSA-5 and HSA-6 hybridoma cell lines were subcloned at the University of Virginia Lymphocyte Culture Center and the cell supernatants were rescreened for activity.

Cloning of the span-x cDNA

A human testis λZAP cDNA library (Freemerman, A. J., Wright, R. M., Flickinger, C. J., and Herr, J. C. (1993). Cloning and sequencing of baboon and cynomologus monkey intraacrosomal protein SP-10: homology with human SP-10 and a mouse sperm antigen (MSA-63). *Mol. Reprod. Dev.* 34, 140-148) was probed with the S71 and S72 mAbs. Using *E. coli* XL-1 Blue as host bacterium, the library was plated at a concentration of 5×10^4 plaque-forming units on NZY plates. After 3.5 hour growth at 42°C, the cells were induced with 10 mM isopropyl-β-D-thiogalactoside (IPTG) on nitrocellulose filters (Micron Separations Inc., Westboro, MA) and incubated at 37°C for 3.5 hours. To block non-specific protein binding sites; the nitrocellulose filters were preincubated with PBS containing 5% non-fat dry milk, 1% BSA, 1% normal goat serum (NGS), 3% Tween-20,

0.15% gelatin and 0.05% thimerosal. Duplicate filters were incubated with the mAbs S71 and S72 individually followed by horseradish peroxidase (HRP)-conjugated goat anti-mouse IgG+IgM. The color reaction was developed with 0.04% diaminobenzidine and 0.03% H_2O_2. Clones were selected on the basis of recognition by both mAbs. *In vivo* excision of the pBluescript phagemid and production of the cDNA-bearing plasmid was performed according to the Stratagene protocol (Stratagene, La Jolla, CA). The longest clone, designated SPANXa, was sequenced.

A human testis λDR2 cDNA library (Clontech, Palo Alto, CA) was probed with the SPAN-Xa cDNA probe. A SPAN-Xa cDNA probe was labeled with digoxigenin (DIG) by PCR using the Genius Nonradioactive Detection System (Boehringer Mannheim, Indianapolis, IN) and SPAN-Xa-specific primers. The λDR2 library was plated and screened according to the manufacturers protocols (Clontech). Filters were hybridized with the DIG-labeled SPAN-Xa cDNA probe and washed to a final stringency of 0.2X SSC (20-strength SSC: 3 M NaCl, 0.3 M sodium citrate-H_20, pH 7.0) and 0.1% SDS at 65 °C. The membranes were then blocked, incubated with alkaline phosphatase (AP)-conjugated anti-DIG antibodies, washed and incubated with a chemiluminescent AP substrate according to the manufacturer's protocols (Boehringer Mannheim). Positive clones were selected by aligning culture plates with developed films, amplified and rescreened twice. *In vivo* excision of the λDR2 phagemid and conversion to the cDNA-containing plasmid pDR2 was performed according to the Clontech protocol.

Plasmid DNA was purified utilizing the Qiagen plasmid mini-prep kit (Qiagen, Valencia, CA). Plasmids were sequenced in both directions by four color automated sequencing at the University of Virginia Biomolecular Research Facility. Sequence analysis was performed using the GCG software (Genetics Computer Group, Madison, WI),

Northern analysis

Northern analysis was performed on two human multiple tissue Northern blots and a human RNA master blot (Clontech). A SPAN-Xa cDNA probe was labeled with DIG by PCR using the Genius Nonradioactive Detection System (Boehringer Mannheim) and SPAN-Xa-specific primers. The membranes were hybridized with the probe corresponding to nucleotides 7-304. Following hybridization, the membranes were washed to a final stringency of 0.2X SSC (20-strength SSC: 3M NaCl, 0.3 M sodium citrate-H_20, pH 7.0) and 0.1% SOS at 65°C. The membranes were then blocked, incubated with AP-conjugated anti-DIG antibodies, washed and incubated with a chemiluminescent AP substrate according to the manufacturers protocols. Chemiluminescent bands and spots were

visualized by autoradiography. The membranes were stripped and probed with a PCR-generated DIG-labeled β-actin cDNA probe as a positive control.

In Situ Hybridization to Human Testis Sections

To generate riboprobes for *in situ* hybridization SPAN-Xa-specific PCR primers were designed containing T7 and SP6 sequences and used to amplify SPAN-Xa cDNA terminating in T7 (5') and SP6 (3') sequences (Birk, P. E., and Grimm, P. C. (1994). Rapid nonradioactive in situ hybridization for interleukin-2 mRNA with riboprobes generated using the polymerase chain reaction. *J. Immunol. Methods.* 167, 83-89., J Immunol Methods 167:83-89). The 5' primer used was:
5'-TAATACGACTCACTATAGGGAGAAAGAGGAGCGTCCCCTGTGATT-3'
(SEQ ID NO:11) and included SPAN-Xa nucleotides 67-88. The 3' primer used was:
5'-GATTTAGGTGACACTATAGMTACTTCCATGAATTCCTCCTCCTCC-3'
(SEQ ID NO:12) and contained the reverse complement of SPAN-Xa nucleotides 282-303. Tritiated UTP was incorporated into the riboprobes by either T7 (for sense strand riboprobes) or SP6 (for antisense riboprobes) RNA polymerase using SPAN-Xa cDNA as a template. A labeled β-actin riboprobe was used as a positive control.

Testes were obtained from patients undergoing elective orchiectomies. Testes were sliced once with a razor blade and immersed in neutral buffered formalin (4%) solution (Sigma) for one hour. The tissue was then minced and placed into fresh fixative overnight. The tissue was dehydrated in a graded series of ethanols, cleared in xylene, and embedded in paraffin. Two and one-half (2.5) μm thick sections were cut, mounted onto 3-aminopropyl triethoxysilane-coated slides, de-paraffinized, rehydrated and treated with Proteinase K (Stoler, M. H., C. R. Rhodes, A. Whitbeck, S. V. Wolinsky, L. T. Chow, and T. R. Broker (1992) Human papillomavirus 16 and 18 gene expression in cervical neoplasias. *Human Pathol.* 23, 117-128). The *in situ* hybridization solution contained 50% formamide, 0.3 M NaCl, 20 mM Tris-HCl, 1 mM EDTA, pH 8.0, 1X Denhardt's solution, 500 μg/ml yeast tRNA and 10% dextran sulfate. The final probe concentration was normalized for probe length and applied at full saturation (0.2 μg/ml/kb complexity). Following hybridization, the sections were washed under high stringency conditions to remove non-specific hybridization. The slides were overlayed with autoradiography emulsion, exposed for 2-4 weeks at 4°C, developed photographically, and lightly stained with hematoxylin-eosin.

Southern Analysis of Human-Mouse Somatic Cell Hybrids

Human/mouse somatic cell hybrids employed for chromosome mapping of human genes had been previously isolated and characterized (Shows, T. B., Sakaguchi, A. Y., and Naylor, S. L. (1982). Mapping the human genome, cloned genes, DNA polymorphisms, and inherited disease. In "Advances in Human Genetics" (H. Harris, and K. Hirschhorn, Eds.), Vol. 12, pp. 341-452. Plenum Press, New York). The mapping panel consisted of 33 somatic cell hybrids derived from 17 unrelated human cell lines, 4 different mouse cell lines, and a Chinese hamster line. The hybrids were previously characterized by chromosomal analysis, chromosome-specific enzyme markers, and in some cases, mapped DNA probes. DNA was isolated from human, mouse, and human/mouse hybrid cells for Southern blot analysis. Ten µg of DNA from each cell line was digested with the restriction enzyme *EcoRI* and fractionated on a 0.8% agarose gel. Fractionated DNA was transferred to nitrocellulose filters and hybridized with the DIG-labeled SPAN-Xa cDNA probe as described above. The DNA used as a probe for hybridization was a PCR-generated DIG-labeled 169-bp fragment of SPAN-Xa cDNA corresponding to nucleotides 67-236. Following hybridization, the membrane was washed to a final stringency of 0.2X SSC and 0.1% SOS at 65C. The membrane was then blocked, incubated with AP-conjugated anti-DIG antibodies, washed and incubated with a chemiluminescent AP substrate according to the manufacturer's protocols (Boehringer Mannheim). Chemiluminescent bands were visualized by autoradiography.

A table was compiled from the 33 cell hybrids by scoring the presence (+) or absence (-) of human bands in the hybrids on the Southern blots. The scoring was related to the presence or absence of specific human chromosomes in each hybrid. A 0% discordancy indicated a matched segregation of the DNA probe with a specific human chromosome.

Fluorescence In Situ Hybridization (FISH)

The forward and reverse primers used for PCR amplification of human genomic DNA correspond to nucleotides 4-25 and 282-303 of the SPAN-Xa cDNA. The 1100 bp-genomic fragment was electroblotted and hybridized with a DIG-labeled SPAN-Xa cDNA probe to confirm that it contained SPAN-Xa DNA sequence. The genomic fragment was then labeled with DIG by PCR for use as a probe for FISH (Boehringer Mannheim).

FISH was performed on XY male lymphocyte metaphase chromosomes as described by Golden, W. L., von Kap-Herr, C., Kurth, B., Wright, R. M., Flickinger, C. J., Eddy, R., Shows, T., and Herr, J. C. (1993). Refinement of the localization of the gene for human intraacrosomal protein SP-10 (ACRV1) to the junction of bands q23-q24 of

chromosome 1·1 by nonisotopic in situ hybridization. *Genomics.* 18, 446-449, and by Matsuda, Y., and Chapman, V. M. (1995). Application of fluorescence in situ hybridization in genome analysis of the mouse. *Electrophoresis.* 16, 261-272.. Slides containing banded metaphase chromosomes were pre-treated with RNase and proteinase K, acetylated to block non-specific hybndization, and dehydrated through an ethanol series. The chromosomal DNA was denatured in 70% formamide, fixed, and hybridized with a denatured DIG-labeled genomic probe. Following stringent washes, the slides were blocked and incubated with FITC-conjugated anti-DIG Fab antibodies. A second layer of chromophore was added by incubation with FITC-conjugated anti-Fabs. Control slides for non-specific hybridization were prepared in the same manner as the experimental slides. Chromosomes were counterstained with 0.1 μg/ml DAPI and mounted. Metaphase chromosomes were viewed with an Olympus Vanox-T AHI2 photomicroscope equipped with epifluorescence optics. DAPI-banding was viewed with a 420-nm barrier filter and specific F1TC signals denoting hybridization were visualized with a 515-nm barrier filter. Metaphase spreads were analyzed and chromosomes were scored as positive for hybridization when both sister chromatids display aligned FITC signals. Specific hybridization signals were recorded on a chromosome histogram.

Construction of expression cassettes

For expression as a recombinant protein, the SPAN-Xa open reading frame (ORE) was placed under the control of the bacteriophage T7 RNA polymerase/promoter system of the *E. coli* expression vectors pET22b(+) and pET28 (Novagen, Milwaukee, WI). To generate the pET22/SPAN-Xa cDNA construct, the 5' primer used was 5'-CATGCCATGGAAAGCCTGCCACTGACATTG-3' (SEQ ID NO:13) containing nucleotides 1-19 and the 3' primer was 5'-ATAGT1TAGCGGCCGCC1TUGCAGGTATTUCAACC-3' (SEQ ID NO:12) containing nucleotides 309-327. The 5' primer used for generating the pET28/SPAN-Xa cDNA construct was 5'-CATGCCATGGACAAACAATCCAGTGCCGGC-3' (SEQ ID NO:14) containing nucleotides 37-60. The 3' primer used for generating the pET28 construct was the same as above. The 5' and 3' primers carried a *NcoI* and a *NotI* restriction site, respectively, in the pET22/SPAN-Xa 5' primer, an additional 'A' was inserted between the Ncol site and the first codon of the cDNA to maintain the reading frame with respect to the vector and, in the 3' primer, an additional 'C' was inserted between the *NotI* site and the cDNA. The 3' primer was designed such that the in-frame termination codon was not included to allow read through into the His_6-tag region on the vector. Following PCR of the SPAN-Xa cDNA, the band was gel purified, cleaved with *NcoI* and *NotI,* and ligated to

NcoI/NotI-digested pET22b(+) or pET28 vector DNA. The plasmid constructs pET22/SPAN-Xa and pET28/SPAN-Xa were verified by automated sequencing. The constructs were each transformed into the *E. coli* expression host, BL21 (DE3), containing a chromosomal copy of the T7 RNA polymerase gene under the control of IPTG-inducible lacUV5 promoter. The *E. coli* BL21 (DE3) strain bearing pET22/SPAN-Xa and pET28/SPAN-Xa were referred to as BL21(DE3)[pET22/SPAN-Xa] and BL21(DE3)[pET28/SPAN-Xa], respectively.

Induction and immobilized metal affinity chromatography (IMAC) purification of recombinant protein (recSPAN-X) was performed as previously described with modifications (Reddi, P. P., Castillo, J. R., Klotz, K., Flickinger, C. J., and Herr, J. C. (1994). Production in *Escherichia coli*, purification and immunogenicity of acrosomal protein SP-10, a candidate contraceptive vaccine. *Gene.* 147, 189-195). The IMAC-purified recSPAN-X was dialyzed extensively against PBS and concentrated using a Centricon 10 microconcentrator (Amicon, Beverly, MA). Protein concentrations were determined by using the bicinchoninic acid (BCA) method according to the manufacturer's specifications (Pierce, Rockford, IL). Purified recombinant protein was analyzed by SDS-PAGE and Western blotting with the S71 mAb. The major band migrating at the expected molecular weight (16 kDa) of the recombinant protein was immunoreactive with the S71 mAb.

Preparation of monospecific polyclonal antibodies (pAbs)

To prepare the recombinant protein for mouse immunizations, individual bands were fractionated by preparative SDS-PAGE and visualized by staining with $CuCl_2$ (Lee, C., Levin, A., and Branton, D. (1987). Copper staining: A five-minute protein stain for sodium dodecyl sulfate-polyacrylamide gels. *Anal. Biochem.* 166, 308-312). Copper staining: A five-minute protein stain for sodium dodecyl sulfate-polyacrylamide gels. *Anal. Biochem.* 166, 308-312). Three bands of approximately 14, 16 and 18 kDa were excised from the gel. To confirm purity and immunoreactivity, individual gel slices were electrophoresed and transferred to nitrocellulose for staining with amido black stain or the S71 mAb. For immunization of female mice, the three gel slices were homogenized with Freund's complete adjuvant for primary injections and with Freund's incomplete adjuvant for two booster injections given at two week intervals. The IMAC-purified solubilized recombinant protein was used for immunization of female guinea pigs. Preimmune serum was collected before injection. Immune serum was obtained ten days after the final injection. All animal investigations were conducted in accordance with the *Guide for Care and Use of laboratory Animals*.

Sperm Preparation and Extraction

Semen specimens were donated by normal, healthy young men. Only ejaculates with normal semen parameters (World Health Organization. (1992). WHO Laboratory Manual for the Examination of Human semen and Semen-Cervical Mucus Interaction. Cambridge University Press, Cambridge, UK) were used in this study. All samples were obtained under informed consent using forms approved by the University of Virginia Human Investigation Committee. Individual semen samples were allowed to liquify at room temperature and washed by centrifugation. For some experiments, mature spermatozoa were separated from seminal plasma, immature germ cells and non-sperm contaminating cells (mainly white blood cells and epithelial cells) by the swim-up technique. For some experiments, mature spermatozoa were separated by Percoll (Pharmacia Biotech, Piscataway, NJ) density gradient centrifugation. For two-dimensional electrophoresis, human sperm pellets were prepared as described by Naaby-Hansen, S., Flickinger, C. J., and Herr, J. C. (1997). Two-dimensional electrophoretic analysis of vectorially labelled surface proteins of human spermatozoa. *Biol. Reprod.* 56, 771-787.

To assess relative protein solubility, washed spermatozoa were extracted in 0.55 CHAPS in TBS (10 mM Tris-HCl, pH 7.5, 150 mM NaCl and protease inhibitor cocktail [Roche Molecular Biochemicals, Indianapolis, IN], with or without 600 mM KCl, for 2 hours at 4°C. In some experiments, 2 mM DTT was added to the extraction buffer. The suspension was then centrifuged at 10,000 x g for 20 minutes. Both the supernatant fluid and pellet fractions were utilized for SDS-AGE and Western blotting.

Protein concentrations were determined by using the Pierce bicinchoninic acid (BCA) method according to the manufacturer's specifications, employing bovine serum albumin (BSA) as a standard.

Polyacrylamide Gel Electrophoresis and Western Blotting

For one-dimensional SDS-PAGE, electrophoresis was performed on 15% acrylamide gels (Laemmli, U. K. (1970). Cleavage of structural proteins during the assembly of the head of bacteriophage T_4. *Nature.* 227, 680-685) with 50 pg of protein per lane. Following SDS-PAGE, polypeptides were either visualized by amido black staining or transferred onto nitrocellulose (Towbin, H., Staehelin, T., and Gordon, J. (1979). Electrophoretic transfer of proteins from polyacrylamide gels to nitrocellulose sheets: procedures and some applications. *Proc. Natl. Acad. Sci. USA.* 76, 4350-4354). Non-reducing gels omitted β-mercaptoethanol or dithiothreitol from the sample buffer. For two-dimensional electrophoresis, isoelectnc focusing (IEF), SDS-PAGE, and electrotransfer was performed as previously described (Naaby-Hansen, S., Flickinger, C. J., and Herr, J. C.

(1997). Two-dimensional electrophoretic analysis of vectorially labelled surface proteins of human spermatozoa. *Biol. Reprod.* 56, 771-787).

Western blots were incubated in PBS containing 0.05% Tween-20 and 5% non-fat dry milk to block nonspecific protein-binding sites. Blots were washed with PBS containing 0.05% Tween-20 between all subsequent incubation steps. They were incubated in S71 mAb, null ascites, pre-immune sera, post-immune sera or affinity-purified mouse pAbs (below) diluted in PBS-NGS or PBS-NDS followed by HRP-conjugated $F(Ab)_2$ fragments of goat anti-mouse IgG/IgM or donkey anti-guinea pig IgG, respectively. HRP conjugates were visualized utilizing TMB reagent following the manufacturer's protocol (Kirkegaard & Perry Laboratories, Gaithersburg, MD).

Affinity purification of polyclonal antisera

Blot affinity purification of the post-immune mouse sera was performed according to the methods of Olmsted, J. B. (1986). Analysis of cytoskeletal structures using blot-affinity purified monospecific antibodies. *Methods Enzymol.* 134, 467-472. with modifications as described (Westbrook-Case, V. A., Winfrey, V. P., and Olson, G. E. (1994). Characterization of two antigenically related integral membrane proteins of the guinea pig sperm periacrosomal plasma membrane. *Mol. Reprod. Dev.* 39, 309-321). Briefly, human sperm SDS extracts were electrophoresed by preparative SDS-PAGE and proteins were electroblotted onto nitrocellulose. A test strip was cut from the blot and immunostained with the pAbs to identify immunoreactive proteins. The band corresponding to immunoreactive material was excised from the blot, blocked with PBS containing 5% NGS, and incubated with the pAbs overnight at 4°C. The excised blot was washed extensively with PBS-NGS and bound antibodies were eluted by incubation for two minutes at room temperature in 0.2 M glycine-HCl, pH 2.5, 0.5 M NaCl and 0.1% Tween 20. The eluant was removed, buffered to pH 8.0 with 1 mM Tris, and dialyzed against PBS. The blot affinity-purified antibodies were evaluated by immunoblot analysis of human sperm extracts to confirm activity.

Immunofluorescence Microscopy

Swim-up spermatozoa were fixed at 4°C with 2% formaldehyde (Electron Microscopy Sciences, Ft. Washington, PA) in PBS for 20 minutes. In some experiments, the cells were not fixed with formaldehyde. Fixed or non-fixed cells were air-dried onto slides and washed three times with PBS. To permeabilize the cells, the slides were incubated in methanol and washed with PBS. Non-specific protein binding sites were blocked by incubating the slides in PBS with 10% NGS or normal donkey serum (NDS).

Slides were incubated with either pre-immune mouse or guinea pig serum (1:250), immune mouse or guinea pig serum (1:250), or blot affinity-purified mouse antibodies (neat) diluted in PBS with 1 % NGS or NDS (PBS-NS). The slides were washed and incubated with FITC-conjugated $F(Ab)_2$ fragments of goat anti-mouse IgG/IgM (1:200) or donkey anti-guinea pig IgG (1:200) in PBS-NS (Jackson ImmunoResearch, West Grove, PA). For confocal microscopy, sperm DNA was stained with propidium iodide during incubation with the secondary antibody. Slides were washed with PBS and mounted with Slow-Fade Light (Molecular Probes, Eugene, Oregon). Cells were observed by differential interference contrast (DIC) and epifluorescence microscopy using a Zeiss axiovert microscope. For confocal microscopy, cells were observed using a Zeiss axiovert microscope. Digital images were obtained using Zeiss LSM software. The proportion of cells containing positive immunostaining was estimated and the 95% confidence interval of the proportion was calculated (Glantz, S. A. (1992). "Primer of Bio-Statistics", 3rd ed. McGraw-Hill, New York).

Combined FISH and Immunoflourescent microscopy of human spermatozoa

Fluorescence in situ hybridization using a green fluorochrome labeled X chromosome probe was performed on swim-up spermatozoa according to the manufacturer s protocols (Vysis, Downers Grove, IL) with modifications. Spermatozoa were air-dried onto slides, fixed with methanol: glacial acetic acid (3:1), and dehydrated through an ethanol series. The sperm nuclei were swollen by incubation in swelling buffer 1 (0.1 M Tris-HCl, pH 8.0, and 10 mM DTT) for 30 minutes followed by incubation in swelling buffer 2 (50 mM Tris-HCl, pH 8.0, and 10 mM LIS: 3,5-Diiodosalicylic acid, lithium salt [Sigma]) for 1-3 hours. Denaturation, hybridization and post-hybridization stringency washes were performed according to the Vysis protocol. The CEP X SpectrumGreen (alpha satellite) was used as a probe for the X chromosome.

Indirect immunofluorescent labeling was performed as above except that a TRITC (rhodamine)-conjugated $F(Ab)_2$ donkey anti-guinea pig IgG (Jackson ImmunoResearch) was used as the secondary antibody. Slides were mounted with Slow Fade containing DAPI II counterstain (Vysis). Cells were observed by epifluorescence microscopy using a Zeiss microscope. Individual blue, green and red fluorescent images were obtained using a digital camera (Hamamatsu) and compiled using Openlab software (Improvision Inc., Boston, MA). The proportions of cells containing red, green, and red/green staining were estimated and the 95% confidence interval of the proportion was calculated (Glantz, S. A. (1992). "Primer of Bio-Statistics", 3rd ed. McGraw-Hill, New York.).

Immunoelectron Microscopy

For post-embedding immunolabeling, washed human spermatozoa were fixed on ice with 4% formaldehyde, 0.5% glutaraldehyde in 0.1 M sodium phosphate buffer, pH 7.4, rinsed in buffer, dehydrated through an ethanol series and embedded in Lowicryl K4M resin (Electron Microscopy Sciences, Ft. Washington, PA). Thin sections were mounted on nickel grids and immunostained. Primary antibodies were used at a concentration of 1:50 (mouse antiserum) or 1:200 (guinea pig antiserum). 5 nm gold-conjugated secondary antibodies (Goldmark, Phillipsburg, NJ) were employed at a concentration of 1:50. Gold-conjugated antibodies used included $F(Ab)_2$ fragments of goat anti-mouse IgG and whole IgG of goat anti-guinea pig IgG. Grids were rinsed with PBS, rinsed with water, stained with uranyl acetate and carbon coated.

6.2. RESULTS

6.2.1. THE S71 AND S72 MABS RECOGNIZE IDENTICAL POLYPEPTIDES

Because the S71 and S72 mAbs were shown to inhibit sperm function (Anderson, D. J. (1990). World Health Organization sperm antigen workshops and sperm antigen project. In "Gamete Interaction: Prospects for Immunocontraception" (N. J. Alexander, D. Griffin, J. M. Spieler, and G. M. H. Waites, Eds.), pp. 103-109. Wiley-Liss, Inc., New York), Western blot analysis was performed to identify their cognate sperm antigens. Both the S71 and S72 mAbs stained identical polypeptide bands on Western blots of human sperm extracts suggesting that these two mAbs recognize identical human sperm proteins (FIG. 2). Under reducing conditions, major immunoreactive bands were clustered into two groups of polymorphic bands: a lower molecular weight (LMW) group of approximately 17, 19, 21, 22 kDa and a higher molecular weight (HMW) group of 33, 34, and 37 kDa. Following subcloning of the cell lines producing the S71 and S72 mAbs, the hybridoma supernatants recognized the identical LMW and HMW groups of proteins on immunoblots (data not shown). Under non-reducing electrophoresis conditions, sperm proteins migrating between 32-35 kDa were immunostained with the S71 and S72 mAbs.

6.2.2. CLONING OF SPAN-X cDNAS

A human testis lZAP cDNA library was probed with S71 and S72 mAbs and clones were selected on the basis of being recognized by both mAbs. The largest clone, designated SPAN-Xa (Genbank accession number AF098306), was plaque purified and sequenced (FIG. 3A). Rescreening this library with radiolabeled SPAN-Xa cDNA probe did not identify longer clones. The 386 bp SPAN-Xa cDNA contained an ORF of 291 bp (97 amino acids) encoding a protein with an expected molecular weight of 11,038.4 Daltons

and isoelectric point (pI) of 4.88 (FIG. 3B). Analysis of the 5' end of the *SPANX* gene obtained from PAC433M19 reveals an in-frame stop codon 5' to this ATG and no other intervening initiator codons supporting the assignment of the initiator methionine to position 37-39 of the SPAN-Xa cDNA. This putative initiator methionine (bold ATG) is surrounded by a sequence that is conserved Kozak, M. (1984). Compilation and analysis of sequences upstream from the translational start site in eukaryotic mRNAs. *Nucleic Acids Res*. 12, 857-872, consensus sequence (AxxxACa**ATG**G: capital letter indicates match with most highly represented nucleotide in that position, small letter indicates match with second highest conserved nucleotide, and x indicates match with poorly conserved nucleotide; Kozak, M. (1984). Compilation and analysis of sequences upstream from the translational start site in eukaryotic mRNAs. *Nucleic Acids Res*. 12, 857-872. Compilation and analysis of sequences upstream from the translational start site in eukaryotic mRNAs. *Nucleic Acids Res*. 12, 857-872). The 72 bp 3' untranslated region contained a consensus polyadenylation sequence (AAUAAA; italics) 34 nucleotides downstream of the TAG termination codon (bold) and 13 nucleotides upstream of the poly A^+ tail.

A second clone, designated SPAN-Xb (Genbank accession number AF098307), was obtained by screening a human testis lDR2 cDNA library (Clontech) with a DIG-labeled SPAN-Xa cDNA probe (FIG. 3A). The 472 bp SPAN-Xb cDNA contained an ORF of 309 bp (103 amino acids; FIG. 3B) encoding a protein with an expected molecular weight of 11826.0 Daltons and a pI of 6.05.The SPAN-Xa and SPAN-Xb cDNA sequences were 92.2% identical. The SPAN-Xb cDNA contained an additional 18 nucleotides encoding 6 in-frame amino acids. The first 9 bp of the 18 bp insert was highly homologous to the previous two sets of 9 nucleotides and may represent a duplication (FIG. 4B, underline). Each group of 9 nucleotides is underlined in FIG. 4B. The peptide sequences encoded by the SPAN-Xa and SPAN-Xb cDNAs were 83.5% identical (FIG. 3B).

Computer analysis revealed similarity of the SPAN-Xa and SPAN-Xb cDNAs with expressed sequence tags (ESTs: AA412605, AI208372, AA412270, AI808260, AI143898, AA382423, AA382424; FIG. 4b) obtained from human testes and two genomic clones. The homologous genomic sequences included a human DNA sequence from clone 376H23 on the X chromosome (accession #AL031078) and a P1 artificial chromosome (PAC) clone 433M19 on chromosome Xq26.3-Xq27.1 (accession #Z95703). The additional 18 nucleotides observed in the SPAN-Xb cDNA were present in the EST95629 clone (AA382424) and the genomic sequence of clone 376H23 (AL031078) at an intron-exon junction (FIG. 4A).

The SPAN-Xa cDNA clone nucleotides 1-108, corresponding to exon 1, were 99.1% identical to PAC 433M19 (Z95703) reverse complement (rc) base pairs 42537-

42644 and the SPAN-Xa nucleotides 109-384 (exon 2) were 96.4% identical to the rc nucleotides 43292-43567. The 648 bp of intervening intronic sequence contained consensus donor and acceptor nucleotides for splice site junctions. The 18 bp insertion observed in the SPAN-Xb cDNA and EST95629 (AA382424) clones was not identified as additional exon sequences in the PAC43319 genomic clone either 5' of the first exon nor within the intron sequences (FIG. 4 A and B). However, this insertion was observed in the genomic DNA sequence of clone 376H23 (AL031078). The SPAN-Xb nucleotides 1-193 corresponding to exon 1 were 100% identical with the 376H23 (AL031078) clone nucleotides 139091-139283 including the 18 bp insertion. The SPAN-Xb nucleotides 194-469 (exon 2) were 100% identical to 376H23 (AL031078) clone nucleotides 139931-140206. The 648 bp intron sequence contained consensus donor and acceptor nucleotides for splice site junctions and was 99.4% identical to the PAC433M19 648 bp intron sequence. These results indicated that the *SPANX* gene is comprised of two exons encoding the cDNA sequence(s) with 648 ~~bp of intervening~~ intron sequence (FIG. 4 A and B). The first exon of the 376H23 genomic sequence contains an 18 bp insertion that is not present in the PAC433M19 genomic sequence although the intron sequence, as well as sequences 5' and 3' of the exons, are highly homologous. This variation in the two genomic sequences indicates genetic polymorphism in the *SPANX* gene within the human population.

Hydropathy analysis of the SPAN-Xa and SPAN-Xb ORFs identified largely hydrophilic peptide sequences (FIG. 5). Hydrophobic regions indicative of transmembrane domains were not identified. The deduced amino acid sequences of SPAN-Xa and SPAN-Xb contained an unusually large percentage of charged amino acid residues. The total number of negatively charged residues in the SPAN-Xa peptide sequence was 18 (18.6%) and positively charged residues was 15 (15.5%). The total number of negatively and positively charged residues in the SPAN-Xb ORF was 19 (18.4%) and 19 (18.4%), respectively. Charged residues, comprising 34% and 37% of the SPAN-Xa and SPAN-Xb total polypeptide, respectively, were evenly dispersed throughout the peptide sequence although there is some grouping of the acidic residues in the N-terminal third, basic residues in the middle third, and acidic residues in the C-terminal third. The significance of this arrangement, if any, is unknown.

Motif analysis identified two overlapping consensus nuclear localization signals (NLS; Hicks, G. R., and Raikhel, N. V. (1995). Protein import into the nucleus: an integrated view. *Annu. Rev. Cell. Dev. Biol.* 11, 155-158) at amino acids 37-43 (PAPKKMK) and 39-45 (PKKMKTS) of the SPAN-Xa ORF (FIG. 2C). These amino acids are conserved in the SPAN-Xb peptide sequence. The SPAN-Xb peptide sequence contains one potential site for N-linked glycosylation at amino acid 28 (within the insertion). The

derived amino acid sequences showed no similarity with known proteins other than those deduced from the ESTs noted above.

6.2.3. SPAN-X IS A TESTIS-SPECIFIC GENE PRODUCT

To determine the size and tissue distribution of the SPAN-X transcript(s), poly(A)$^+$ RNA from 16 human tissues (Clontech) was examined by Northern blotting. A broad band of 0.60 kb was observed exclusively in human testis when the membrane was probed with a DIG-labeled SPAN-Xa cDNA probe (FIG. 6). With less exposure time, several transcripts could be resolved within the broad band (data not shown). This could be due, in part, to genetic polymorphism as described above and/or alternative splice variants. Further investigations will evaluate the relative abundance and sequence(s) of multiple SPAN-X mRNAs. No transcripts were identified in other human tissues including peripheral blood leukocytes, colon, small intestine, ovary, prostate, thymus, and spleen (FIG. 6), as well as, heart, brain, placenta, lung, liver, skeletal muscle, kidney, and pancreas (not shown). Fifty human tissue RNA samples were screened on a RNA master dot blot (Clontech). Again, only the testis RNA sample showed hybridization with the SPAN-Xa cDNA probe (data not shown). To demonstrate that intact mRNA was present in each tissue, the membranes were stripped and probed with a DIG-labeled β-actin cDNA probe. A β-actin signal was observed in all tissues. These results indicate that SPAN-X is a testis-specific gene product.

6.2.4. SPAN-X GENE MAPS TO CHROMOSOME Xq27.1

To define the genetic locus of the *SPANX* gene, DNA samples from a panel of human/mouse somatic cell hybrids were probed with DIG-labeled SPAN-Xa cDNA. Southern analysis demonstrated mapping of the SPAN-Xa cDNA probe to human chromosome X by somatic cell hybrids. There was complete concordance between the presence of the human DNA fragments that hybridized to the SPAN-Xa probe and the presence of chromosome X, i.e. 0% discordancy for chromosome X (Table 1).

Table 1

Chromosome	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	X
Concordant # of hybrids	22	18	16	18	17	21	22	18	13	20	16	19	19	18	19	16	15	14	19	13	15	18	24
Discordant # of hybrids	10	15	15	14	16	12	11	15	18	12	13	14	14	14	11	17	16	19	14	20	18	14	0
% Discordancy	31	45	48	44	48	36	33	45	58	38	45	42	42	44	37	52	52	58	42	61	55	44	0

Discordancy analysis of SPAN-Xa cDNA with human chromosomal markers. The SPAN-Xa cDNA probe was hybridized to Southern blots containing *EcoR*I digested DNA from human/mouse cell hybrids. Concordant hybrids have either retained or lost the human

bands together with a specific human chromosome. Discordant hybrids have either retained the human bands, but not a specific chromosome, or retained the chromosome but not the bands. Percent discordancy indicates the degree of discordant segregation for a marker and a chromosome. A 0% discordance is the basis for chromosome assignment. The table demonstrates mapping of the SPAN-Xa DNA probe to human chromosome X in these somatic cell hybrids (i.e., 0% discordancy for chromosome X).

Hybrid cell lines containing translocations of human chromosome X allowed further refinement of the *SPANX* gene to the Xq22-Xq28 region (Table 2).

Table 2

Hybrid	SPAN-Xa	Translocations
DUA-1A	+	Xqter-Xp11::15q11-15qter
DUA-1CSAZ	-	15pter-15q11::Xp11-Xpter
ATR-13	+	5pter-5q35::Xq22-Xqter
REX-11BSHF	+	22pter-22q13::Xq22-Xqter
XTR-3BSAGB	-	3pter-3q21::Xq28-Xqter
XTR-3BSAGH	-	3pter-3q21::Xq28-Xqter

Southern blot analysis with SPAN-Xa cDNA probe of hybrid cell lines containing X translocations but no intact X chromosome. The presence (+) or absence (-) of SPAN-Xa bands in the hybrid on the blots is indicated. These hybrids allowed further refinement of the SPAN-X locus to the Xq22-Xq28 region.

These results were confirmed by fluorescence in situ hybridization (FISH) of human male metaphase chromosomes with a DIG-labeled SPAN-Xa genomic probe. Hybridization of the probe was observed as a paired fluorescent fluorescein signal on the proximal region Xq27 Of 40 XY male metaphases with paired hybridization signals, 22 metaphases exhibited paired signals on Xq27. No other chromosomal bands, on the autosomes or the Y chromosome, exhibited paired signals greater than once. The band localization was achieved by selecting 10 metaphase spreads hybridized with the genomic probe that displayed the best DAPI banding. On the 10 positive cells examined, all signals localized to the proximal region of Xq27, specifically Xq27.1. A biotin-labeled α satellite probe (DX21) identified by avidin-Texas red staining was used as a marker for the X chromosome. The localization of SPAN-X by FISH to Xq27.1 is consistent with the homology of the SPAN-Xa cDNA clone to the PAC clone 433M19 on chromosome Xq26.3-Xq27.1.

6.2.5. SPAN-X mRNA IS TRANSCRIBED POST-MEIOTICALLY

In situ hybridization was performed to determine the pattern of mRNA expression in human testes (FIG. 7). On sections hybridized with a tritium-labeled SPAN-Xa antisense riboprobe (complementary to endogenous tissue mRNA), silver grains were seen primarily in the most luminal cell types corresponding to round and elongating spermatids (A, B). Elongated spermatids with highly condensed nuclei showed fewer silver grains (C). Few grains were observed on spermatogonia and spermatocytes. Only background staining was observed on testis sections hybridized with the negative control sense probe (D). This indicates that the *SPANX* gene is transcribed post-meiotically.

6.2.6. PRODUCTION AND PURIFICATION OF RECOMBINANT PROTEIN

To produce recombinant protein, the ORF of SPAN-Xa cDNA was subcloned into the *E. coli* expression vectors pET22 and pET28 (Novagen) with the ORF under the control of the T7 bacteriophage RNA polymerase/promoter system. The pET22/SPAN-Xa and pET28/SPAN-Xa constructs were transformed into the *E. coli* expression host BL21 containing a chromosomal copy of the T7 RNA polymerase gene under control of an IPTG-inducible promoter. RecSPAN-X was expressed with a six histidine tag and purified by IMAC.

The pET28/SPAN-Xa expressed, IMAC-purified fraction was electrophoresed by SDS-PAGE and stained with amido black stain or transferred and immunostained with the S71 mAb (FIG. 8). The major band migrating at the expected molecular weight (16 kDa) of the recSPAN-X was immunoreactive with the S71 mAb.

6.2.7. ANTI-SPAN-X ANTIBODIES RECOGNIZE POLYMORPHIC PROTEINS ON WESTERN BLOTS

Four female mice and four female guinea pigs were immunized with the purified recSPAN-X protein to produce pAbs. Human sperm SDS extracts were separated by reducing and non-reducing one dimensional SDS-PAGE and immunostained with pre-immune serum and immune serum from each of the eight immunized animals or with blot affinity-purified mouse antibodies. On Western blots of sperm proteins electrophoresed under reducing conditions, post-immune sera from each immunized animal strongly reacted with a broad band migrating between 15-19 kDa (FIG. 10). These bands were identical to the LMW proteins recognized by the S71 mAb. The HWM proteins recognized by S71 were not reactive with any of the eight pAbs generated against recSPAN-X. Blot affinity-purified mouse antibodies showed identical staining pattern on Western blots of human

sperm SDS extracts (data not shown). Under non-reducing SDS-PAGE of sperm SDS extracts, polymorphic immunoreactive bands were visualized at a similar molecular weight to the LMW proteins seen in reduced sperm extracts. In addition, a group of immunoreactive polymorphic proteins was observed between approximately 28-35 kDa. Pre-immune sera showed only faint background immunoreactivity with sperm extracts.

These data confirmed that the SPAN-X cDNAs cloned from the human testis libraries encode the LMW protein recognized by the S71 and S72 mAbs. This was verified by reactivity of the S71 and S72 mAbs with recSPAN-X protein (FIG. 8) and by the co-migration of immunoreactive bands recognized by the S71/S72 mAbs and by the pAbs raised against recSPAN-X on Western blots of sperm extracts (FIG. 9).

To assess the relative solubility of the SPAN-X protein, washed spermatozoa were extracted in a variety of solubilization buffers (FIG. 10). The soluble supernatant fraction and insoluble pellet fraction were examined by immunoblotting with SPAN-X antibodies. SPAN-X was observed in the insoluble pellet fraction of spermatozoa extracted with 0.5% CHAPS, 0.5% CHAPS containing 600 mM KCl, and 0.5% CHAPS containing 600 mM KCl and 2 mM DTT. SPAN-X was observed in the soluble fraction following extraction with 1% SDS. This data indicates that SPAN-X is a relatively insoluble sperm protein.

For two dimensional gel analysis of the SPAN-X protein, Percoll-purified human spermatozoa were extracted in lysis buffer. Sperm extracts were separated in the first dimension by IEF, separated in the second dimension by SDS-PAGE and silver stained or transferred to nitrocellulose and immunostained with the pAbs (FIG. 11). Approximately 19 immunoreactive spots were visualized at approximately 19.8-23.5 kDa with pIs ranging from 5.0-5.5 (FIG. 11: b, b'; Table 3). The net charges of the SPAN-Xa and SPAN-Xb deduced amino acid sequences are 4.88 and 6.05, respectively, that are similar to the pI values observed on two-dimensional immunoblots. On corresponding silver-stained gels, these proteins appeared yellow in color (FIG. 11: a, a'). These results indicate that SPAN-X is a relatively acidic, agyrophobic, polymorphic protein.

Table 3.

Spot #	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
pI	4.99	5.02	5.00	5.03	5.06	5.09	5.06	5.09	5.30	5.24	5.20	5.24	5.30	5.26	5.23	5.28	5.54	5.53	5.44
MW	22.8	21.9	20.9	23.2	22.5	21.3	20.6	23.5	23.0	22.4	22.1	21.8	21.6	20.1	20.3	20.2	21.6	21.3	19.8

Relative mobility (kDa) and pI values of the 19 forms of human sperm SPAN-X as observed on 2D immunoblots.

6.2.8. SPAN-X PROTEIN LOCALIZES TO THE NUCLEAR VACUOLES AND REDUNDANT NUCLEAR MEMBRANE

To localize the SPAN-X antigen on human spermatozoa, indirect immunofluorescence was performed using the anti-recombinant pAbs and control pre-immune sera. Immunofluorescence localization demonstrated intense staining in the nuclear craters (FIG. 12, large arrows; FIG. 13: a, b) and/or cytoplasmic droplets (FIG. 12, small arrows; FIG. 13: c, d, e, f) of formaldehyde-fixed, methanol-permeabilized, swim-up spermatozoa with each post-immune antisera as well as the blot affinity-purified antibodies. SPAN-X staining revealed several phenotypes including small and large nuclear craters (FIG. 13: a, b, respectively); multiple craters (FIG. 13: a); small and large cytoplasmic droplets (FIG. 13: c, d, respectively) without crater staining; and both cytoplasmic droplets and nuclear craters (FIG. 13: e, f). Identical localization was observed on paraformaldehyde-fixed or non-fixed, air-dried spermatozoa indicating that the localization is not an artifact of aldehyde fixation (data not shown). Immunofluorescent staining was not observed on every spermatozoon in the field of view although nuclear craters and/or cytoplasmic droplets were present in these cells (FIG. 12, circled area). Spermatozoa incubated with preimmune sera exhibited no fluorescence (FIG. 12, right panels).

The incidence of SPAN-X immunofluorescent staining was categorized according to SPAN-X phenotypes (Table 4). Using DIC imaging, greater than 95% of spermatozoa exhibited one or more nuclear craters using DIC imaging. Of 1281 spermatozoa examined from a pool of 11 donors by indirect immunofluorescence for SPAN-X protein, 20.4% of spermatozoa exhibited one or more immunofluorescent craters without cytoplasmic droplet staining (FIG. 13: a, b). 25.7% of spermatozoa showed staining of the cytoplasmic droplet without crater staining (FIG. 13: c, d), and 4.5% showed both crater and cytoplasmic droplet staining (FIG. 13: e, f). In total, the SPAN-X protein localized to either nuclear craters, cytoplasmic droplets or both in 50.6% (with a 95% confidence interval of ±2.74%) of spermatozoa by indirect immunofluorescence using conventional microscopy (Table 4).

Table 4.

IIF	+ Crater (1 or more)	261/1281	20.4%
	+CD	329/1281	25.7%
	+Crater and CD	58/1281	4.5%
		648/1281	50.6%
CFM		515/1039	49.6%
Total		1163/2320	**50.1%**

Quantitative analysis of SPAN-X immunofluorescent labeling patterns on formaldehyde-fixed, methanol-permeabilized swim-up spermatozoa by conventional (IIF) and confocal (CFM) microscopy. More than 2000 spermatozoa were scored. In total, 50.1% of spermatozoa demonstrated SPAN-X immunoreactivity.

Because the finding of SPAN-X in only half of the population was a unique observation, fluorescent-labeled, swim-up spermatozoa from a pool of 15 donors were analyzed by confocal microscopy through all planes of the sperm head to further examine the incidence of SPAN-X staining. Sperm were immunolabeled with an FITC-conjugated secondary antibody and the sperm DNA was stained with propidium iodide. 0.5 mm z-sections were taken through the depth of the sperm head by extended depth of focus. Of 1039 sperm examined by digital imaging, 515 (49.6 ±3.0%) exhibited immunofluorescent staining with the anti-SPAN-X antibody (Table 4). In total, 2320 spermatozoa were examined by either conventional or confocal microscopy. Of these 2320 spermatozoa, 1163 (50.1 ±2.0%) exhibited immunofluorescent labeling with the SPAN-X antisera (Table 4).

To determine the number of X- and Y-bearing spermatozoa that contained SPAN-X protein, FISH to identify the X chromosome and immunofluorescence to detect SPAN-X was performed consecutively on swim-up spermatozoa with swollen nuclei. The number of cells exhibiting SPAN-X and X chromosome staining, SPAN-X without X staining, no SPAN-X with X staining, and neither SPAN-X nor X staining was counted, the proportions calculated, and the 95% confidence interval determined for each subset (Table 5). In total, 389 cells were counted and scored. Of these, 24.9% exhibited SPAN-X and X chromosome staining while 27.2% exhibited SPAN-X without X staining. These data indicate that SPAN-X protein was equally distriuted between X- and Y-bearing spermatozoa.

Table 5.

	Green (X)	Not Green
Red (SPAN-X)	97 (24.9% ± 4.3)	106 (27.2% ± 4.4)
Not Red	97 (24.9 ± 4.3)	89 (22.9% ± 4.2)

Quantitative analysis of the number of cells exhibiting SPAN-X and X chromsome staining (Red and Green), SPAN-X without X staining (Red and Not Green), no SPAN-X with X staining (Not Red and Green), and neither SPAN-X nor X staining (Not Red and Not Green). The number of cells in each subset is indicated along with the proportion of the

total and the 95% confidence interval of the proportion. These data show an equal distribution among groups indicating that SPAN-X is associated with both X- and Y-bearing spermatozoa.

Sperm prepared for post-embedding immunoelectron microscopy exhibited a number of randomly distributed cavities throughout the condensed chromatin of the sperm head (FIG. 14). These cavities, identified as nuclear vacuoles, are non-membrane bound areas filled with amorphous, granular material and are devoid of nuclear chromatin. Sperm immunolabeled with the anti-SPAN-X antibodies showed gold labeling of granular material within nuclear vacuoles (FIG. 14A, B). Gold labeling appeared to be associated with electron dense material within the nuclear vacuoles. Little staining was observed in the surrounding nucleoplasm. In addition, specific staining of folds of the extensive redundant nuclear membrane was observed in the cytoplasmic droplet at the base of the sperm head (FIG. 14A, C). Staining on the redundant nuclear membrane was observed caudal to the boundary of the posterior ring. Spermatozoa labeled with pre-immune sera showed no staining of the nuclear vacuoles, redundant nuclear membrane or other sperm organelles.

6.3. DISCUSSION

6.3.1. *SPAN-X*, MAPPED TO HUMAN CHROMOSOME XQ27.1, IS EXPRESSED IN HAPLOID SPERMATIDS

This study demonstrated testis-specific, post-meiotic expression of the X-linked gene *SPANX*. Analysis of human/mouse somatic cell hybrids containing various complements of human chromosomes and FISH analysis of male metaphase chromosomes assign the human *SPANX* gene to Xq27.1. Northern analysis of 50 human tissues indicated that SPAN-X gene expression is specific to the testis. In situ hybridization of human testes showed that SPAN-X mRNA is present in round and elongating spermatids. The presence of label representing SPAN-X mRNA in spermatids and the scarcity of label in spermatogonia and spermatocytes indicates that the message is post-meiotically transcribed. This is significant as it demonstrates haploid expression of an X-linked gene.

The genes expressed during spermatogenesis can be divided into two major groups: diploid expressed genes of spermatogonia and spermatocytes, and post-meiotic, haploid expressed genes of spermatids (Hecht, N. B. (1990). Regulation of 'haploid expressed genes' in male germ cells. *J. Reprod. Fert.* 88, 679-693; Hoog, C. (1995). Expression of a large number of novel testis-specific genes during spermatogenesis coincides with the functional reorganization of the male germ cell. *Int. J. Dev. Biol.* 39, 719-726; Nayernia,

K., Adham, I., Kremling, H., Reim, K., Schlicker, M., Schluter, G., and Engel, W. (1996). Stage and developmental specific gene expression during mammalian spermatogenesis. *Int. J. Dev. Biol.* 40, 379-383). Since chromosomes segregate during meiosis, spermatids are not genetically equivalent; therefore, post-meiotic expression could lead to functionally different spermatozoa. Perhaps to prevent this occurrence, cytoplasmic bridges connecting sister spermatids may provide a mechanism for the sharing of haploid gene transcripts. In accord with this hypothesis, (Braun, R. E., Behringer, R. R., Peschon, J. J., Brinster, R. L., and Palmiter, R. D. (1989). Genetically haploid spermatids are phenotypically diploid. *Nature*. 337, 373-376) observed sharing of human growth hormone (hGH) mRNA and protein products of a transgene, containing the mouse protamine promoter and hGH coding sequence, that was expressed exclusively in post-meiotic germ cells of mice hemizygous for the transgene. Moreover, in mice containing chromosomally unbalanced gametes, (Caldwell, K. A., and Handel, M. A. (1991). Protamine transcript sharing among postmeiotic spermatids. *Proc. Natl. Acad. Sci. USA.* 88, 2407-2411) observed sharing of protamine (Prm-1) transcripts among spermatids within a syncytium. mRNA sharing may be aided by testis-brain RNA-binding protein (TB-RBP), a single-stranded DNA- and RNA-binding protein that moves from the nucleus to the cytoplasm and through intercellular bridges of male germ cells (Han, J. R., Yiu, G. K. C., and Hecht, N. B. (1995). Testis brain-RNA binding protein (TB-RBP) is a microtubule associated protein that attaches translationally repressed and transported mRNAs to microtubules. *Proc. Natl. Acad. Sci. USA.* 92, 9550-9554; Morales, C. R., Wu, X. Q., and Hecht, N. B. (1998). The DNA/RNA-binding protein, TB-RBP, moves from the nucleus to the cytoplasm and through intercellular bridges in male germ cells. *Dev. Biol.* 201, 113-123).

Most haploid-expressed genes have been mapped to autosomal chromosomes (Golden, W. L., von Kap-Herr, C., Kurth, B., Wright, R. M., Flickinger, C. J., Eddy, R., Shows, T., and Herr, J. C. (1993). Refinement of the localization of the gene for human intraacrosomal protein SP-10 (ACRV1) to the junction of bands q23-q24 of chromosome 11 by nonisotopic in situ hybridization. *Genomics.* 18, 446-449; Nelson, J. E., and Krawetz, S. A. (1993). Linkage of human spermatid-specific basic nuclear protein genes. Definition and evolution of the P1à P2 à TP2 locus. *J. Biol. Chem.* 268, 2932-2936; Tarnasky, H., Gill, D., Murthy, S., Shao, X., Demetrick, D. J., and van der Hoorn, F. A. (1998). A novel testis-specific gene, SPAG4, whose product interacts specifically with outer dense fiber protein ODF27, maps to human chromosome 20q11. 2. *Cytogenet. Cell. Genet.* 81, 65-67). Thus, their corresponding gene products would be expressed in all haploid spermatids although allelic variations could occur. In contrast, post-meiotic expression of sex-linked genes would require mRNA and/or protein sharing among sister spermatids through

intercellular bridges for X- and Y-bearing spermatids to be functionally equivalent. In accord with this, an underrepresentation of transcripts assigned to the X chromosome has been reported in human adult testis (Jones, M. H., Zhang, Y., Tirosvoutis, K. N., Davey, P. M., Webster, A. R., Walsh, D., Spurr, N. K., and Affara, N. A. (1997). Chromosomal assignment of 311 sequences transcribed in human adult testis. *Genomics*. 40, 155-167). This may represent the inactivation of the X chromosome during meiosis and selection against X-linked germ cell transcripts. Thus, the mapping of the testis-specific *SPANX* to the X chromosome is of considerable interest.

Post-meiotic mRNA transcription of few X-linked genes has been described Hendriksen, P. J. M., Hoogerbrugge, J. W., Themmen, A. P. N., Koken, M. H. M., Hoeijmakers, J. H. J., Oostra, B. A., van der Lende, T., and Grootegoed, J. J. (1995). Postmeiotic transcription of X and Y chromosomal genes during spermatogenesis in the mouse. *Dev. Biol.* 170, 730-733; Moss, S. B., VanScoy, H., and Gerton, G. L. (1997). Mapping of a haploid transcribed and translated sperm-specific gene to the mouse X chromosome. *Mamm. Genome.* 8, 37-38; Turner, R. M. O., Johnson, L. R., Haig-Ladewig, L., Gerton, G. L., and Moss, S. B.. (1998). An X-lined gene encodes a major human sperm fibrous sheath protein, hAKAP82. *J. Biol. Chem.* 273, 32135-32141). For example, the X-linked gene for AKAP82, a major fibrous sheath protein of mouse (mAKAP82) and human (hAKAP82) sperm flagella, exhibits both post-meiotic and testis-specific gene expression (Moss, S. B., VanScoy, H., and Gerton, G. L. (1997). Mapping of a haploid transcribed and translated sperm-specific gene to the mouse X chromosome. *Mamm. Genome.* 8, 37-38; Turner, R. M. O., Johnson, L. R., Haig-Ladewig, L., Gerton, G. L., and Moss, S. B.. (1998). An X-lined gene encodes a major human sperm fibrous sheath protein, hAKAP82. *J. Biol. Chem.* 273, 32135-32141). Although transcribed in only half of developing male germ cells, AKAP82 is present in all spermatozoa and the mRNA for AKAP82 contains the conserved sequences for recognition by TB-RBP (Morales, C. R., Wu, X. Q., and Hecht, N. B. (1998). The DNA/RNA-binding protein, TB-RBP, moves from the nucleus to the cytoplasm and through intercellular bridges in male germ cells. *Dev. Biol.* 201, 113-123). Although SPAN-X is observed in 50% of ejaculated spermatozoa, the protein is present in both X- and Y-bearing spermatozoa indicating sharing of the mRNA and/or protein between sister spermatids. As the SPAN-X mRNA does not contain the conserved recognition sequence for TB-RBP, another mechanism(s) may be involved in SPAN-X sharing. Thus, AKAP82 and SPAN-X could be utilized to study the mechanisms of sharing endogenous haploid expressed gene products between X- and Y-bearing spermatids.

6.3.2. SPAN-X LOCALIZES TO SPERM NUCLEAR VACUOLES AND REDUNDANT NUCLEAR MEMBRANES

Using immunoelectron microscopy, SPAN-X was identified within sperm nuclear vacuoles and the redundant nuclear membrane. Antibodies raised against recSPAN-X recognize sperm craters and cytoplasmic droplets in ejaculated spermatozoa by indirect immunofluorescence. Sperm craters are thought to correspond to either indentations on the nuclear surface or to vacuoles within the condensed chromatin of the nucleus. With DIC optics, the determination of craters as nuclear surface or subsurface features is not possible. Using immunoelectron microscopy, SPAN-X was identified within the sperm nuclear vacuoles. This result directly identifies nuclear craters observed by DIC imaging as nuclear vacuoles in mammalian spermatozoa. Furthermore, SPAN-X is the first example of a specific protein localized to nuclear vacuoles of mature spermatozoa of any species.

Based on chemical composition and morphological structures, nuclear vacuoles are thought to be derived from the nucleolus of spermatocytes and spermatids (Czaker, R. (1985). Ultrastructural observations on nucleolar changes during mouse spermatogenesis. *Andrologia*; Dadoune, J. P. and Alfonsi, M. F. 1986. Ultrastructural and cytochemical changes of the head components of human spermatids and spermatozoa. *Gamete Res.* 14, 33-46; Sousa, M., and Carvalheiro, J. (1994). A cytochemical study of the nucleolus and nucleolus-related structures during human spermatogenesis. *Anat. Embryo.* 190, 479-487). The nucleolus, which synthesizes the major components of ribosomes, typically contains two forms of RNA: granular, representing maturing RNP particles; and fibrillar, possibly a precursor to the granules. Within the nuclear vacuole of spermatozoa, fibrils and dense fibrillar structures are often associated with the peripheral condensed chromatin. Cytochemical staining has indicated the presence of RNPs and DNPs in the fibrillar structures of nuclear vacuoles (Dadoune, J. P. and Alfonsi, M. F. 1986. Ultrastructural and cytochemical changes of the head components of human spermatids and spermatozoa. *Gamete Res.* 14, 33-46; Sousa, M., and Carvalheiro, J. (1994). A cytochemical study of the nucleolus and nucleolus-related structures during human spermatogenesis. *Anat. Embryo.* 190, 479-487).

RNPs and DNPs have also been identified in stacks of annulate lamellae in human spermatids (Paniagua, R., Nistal, M., Amat, P., and Rodriguez, M. C. (1987). Presence of basic proteins and ribonucleoproteins in the neck region of human spermatids and spermatozoa. *J. Anat.* 151, 137-142). These lamellae are closely associated with the nuclear membrane during spermatogenesis and with the redundant nuclear membrane during late spermiogenesis (Smith and Berlin, 1977; Sun, C. N., Chew, E. C., and White, H. J. (1977).

Cytoplasmic annulate lamellae and intranuclear membranes in human spermatids and sperm. *Cell Biol. Int. Rep.* 1, 345-351; Bird, D. J., and Seiler, M. W. (1986). Annulate lamellae and single-pore complexes in human spermatogonia. *J. Submicrosc. Cytol.* 18, 823-828). Furthermore, these membrane stacks are often found within the nucleus as the spermatid chromatin condenses (Sun, C. N., Chew, E. C., and White, H. J. (1977). Cytoplasmic annulate lamellae and intranuclear membranes in human spermatids and sperm. *Cell Biol. Int. Rep.* 1, 345-351). The annulate lamellae may represent a specialized form of the endoplasmic reticulum and function in transfer of material from the nucleus to the cytoplasm. SPAN-X antibodies show immunoreactivity with membrane-like structures within nuclear vacuoles as well as extensive redundant nuclear membrane within the cytoplasmic droplet of ejaculated spermatozoa. Whether SPAN-X is a component of the annulate lamellae and/or nuclear membrane of spermatids remains to be determined. Furthermore, it is likely that SPAN-X represents a binding protein to either nucleic acids or basic nuclear proteins based on its localization and the charged nature of the SPAN-X protein. During spermatid elongation, the cytoplasmic droplet and redundant nuclear membrane are formed as a result of reduction in cytoplasmic and nuclear volume. Furthermore, the nuclear pore complexes are relocated to the redundant nuclear membrane during spermiogenesis (Chemes, H. E., Fawcett, D. W., and Dym, M. (1978). Unusual features of the nuclear envelope in human spermatogenic cellls. Anat. Rec. 192, 493-512). Exclusion of SPAN-X from the nucleus into the residual cytoplasm/cytoplasmic droplet during spermiogenesis or epididymal sperm maturation may explain the absence of SPAN-X staining in some nuclei. The co-localization of SPAN-X to both nuclear vacuoles and the redundant nuclear membrane suggests that, in some spermatozoa or during some stage of sperm development, these two compartments are contiguous. During spermiogenesis, continuity between the nuclear vacuoles and the redundant nuclear membrane may result in transport of defunct materials such as mRNAs and/or proteins into the cytoplasmic droplet for disposal. Serial sectioning of human sperm would be helpful to demonstrate this phenomenon.

As a marker for nuclear vacuoles and redundant nuclear membranes, SPAN-X may aid in understanding the function of these structures during spermatogenesis, epididymal maturation, and transport through the female reproductive tract. Furthermore, following fertilization, the potential role of SPAN-X in nuclear decondensation and in formation of the male pronucleus may then be examined.

6.3.3. SPAN-X IS A HIGHLY INSOLUBLE, ACIDIC, POLYMORPHIC SPERM PROTEIN

This study characterized several molecular and biochemical properties of SPAN-X, an acidic, structural protein of the human sperm nucleus. The nuclear localization consensus sequence, identified at the middle of the SPAN-X peptide sequence, correlates with the localization of SPAN-X to the human sperm nucleus. This seven amino acid consensus signal, containing a proline followed within the next six residues by three lysines, has been shown to direct translocation of proteins into the nucleus (Hicks, G. R., and Raikhel, N. V. (1995). Protein import into the nucleus: an integrated view. *Annu. Rev. Cell. Dev. Biol.* 11, 155-158). The SPAN-X nuclear localization signal sequence disclosed herein is, however, different from that dislcsoed by Hick et al. in that the NLS of SPAN contains two embedded NLSS motifds In particular, the novel NLSS of SPAN-X contains the specific nine amino acid sequece of PAPKKMKTS (SEQ ID NO:17). Hicks et al does not disclose or suggest the embedded nucear alocasd ssiffanal sequence of SPAN-X.

Similar to the nuclear transition proteins TP1 and TP2, the SPAN-X protein sequence has a large number of lysines, arginines, and histidines. The transition proteins are small, basic, spermatid-specific proteins that bind DNA and aid in the transition from nuclear histones to protamines during the compaction of spermatid DNA (reviewed by Balhorn, R. (1982). Mammalian protamines: structure and molecular interactions. In "Molecular Biology of Chromosome Function" (K. W. Adolf, Ed.), pp. 366-395. Springer, New York;Hecht, N. B. (1989). Molecular biology of structural chromosomal proteins of the mammalian testis. In "Molecular Biology of Chromosome Function" (K. W. Adolf, Ed.), pp. 396-420. Springer, New York.). Although SPAN-X mRNA expression pattern, molecular weight, and number of acidic amino acids is similar to the transition proteins, SPAN-X, an acidic protein, exhibits no sequence similarities to known spermatid nuclear proteins including the transition proteins, spermatid-specific histones, high mobility group proteins, or protamines. Furthermore, SPAN-X protein is present in ejaculated sperm nuclei whereas transition proteins are replaced by protamines during late spermiogenesis (Hecht, N. B. (1989). Molecular biology of structural chromosomal proteins of the mammalian testis. In "Molecular Biology of Chromosome Function" (K. W. Adolf, Ed.), pp. 396-420. Springer, New York).

The SPAN-X protein is a highly polymorphic protein conserved in the human population as indicated by identical SPAN-X immunoreactive bands in every sperm sample (n>50) tested. SPAN-X was originally identified with the S71 and S72 mAbs that recognize two groups of polymorphic proteins on Western blots. All anti-recSPAN-X antisera

recognize the LMW group of S71/S72 immunoreactive proteins but do not recognize the HMW group.

SPAN-X appears to be a disulfide cross-bridged, multimeric protein. Additional immunoreactive bands migrating between 27-30 kDa SPAN-X were observed in human sperm extracts electrophoresed under non-reducing conditions. Whether the SPAN-X complex is composed of different protein subunits is not yet resolved. As SPAN-X was not solubilized under non-denaturing conditions, the mass and charge of native SPAN-X protein complexes were not determined. Nearest neighbor cross-linking experiments are underway to identify potential associations with other sperm nuclear proteins and/or nucleic acids. Differential extraction of human spermatozoa demonstrates that the SPAN-X protein is insoluble in zwitterionic detergents, high salt, and reducing agents but can be solubilized by denaturing detergents. This insolubility of SPAN-X was not due to inaccessibility of the sperm nucleus to extraction buffers as the majority of SPAN-X is associated with the redundant nuclear membrane of the cytoplasmic droplet. The ultrastructural localization of the insoluble SPAN-X protein suggests that SPAN-X is a structural component of the sperm nuclear membrane or is associated with structural components of the nucleus, possibly the nuclear matrix.

Nineteen variants of SPAN-X were detected on two-dimensional immunoblots over a pI range of 5.0-5.5 and molecular mass of approximately 20-23 kDa. The slight differences in apparent molecular mass observed by one- and two-dimensional SDS-PAGE are consistent with differences in sperm extraction and electrophoresis conditions. The observed pI range is consistent with the theoretical pI of the SPAN-Xa and SPAN-Xb deduced amino acid sequences that contain a large number of negatively charged amino acids (19%). The variations in pI and molecular mass of SPAN-X may be the result of proteolysis, post-translational modifications such as phosphorylation or glycosylation, genetic polymorphism, alternative splicing or a combination of these causes. The polymorphism observed in the *SPANX* gene represents one possible cause of SPAN-X protein heterogeneity.

Interestingly, SPAN-X stained yellow using standard silver staining techniques. Although this could potentially be due to the highly charged nature of SPAN-X, previous literature to support this hypothesis has not been found. Rather, previous data indicates that sialyated glycoproteins stain yellow by this method (Deh, M. E., Dzandu, J. M., Wise, G. E. (1985). Sialoglycoproteins with a high amount of O-glycosidically linked carbohydrate moieties stain yellow with silver in sodium dodecyl sulfate-polyacrylamide gels. *Anal. Biochem.* 150, 166-173). Glycosylation of nuclear proteins has been described in somatic cells Polet, H., and Molnar, J. (1988). Demonstration that some of the nonhistone proteins,

inducible to translocate into the nucleus, are glycosylated. *J. Cell. Physiol.* 135, 47-54). Although there are several potential glycosylation sites in the deduced amino acid sequence of SPAN-X, the presence of carbohydrate moieties on the SPAN-X protein remains to be determined.

6.3.4. DISTRIBUTION OF SPAN-X PROTEIN IN 50% OF X- AND Y-BEARING HUMAN SPERMATOZOA

A remarkable observation was that close to 50% of ejaculated human spermatozoa in a donor pool exhibited immunofluorescent labeling with the SPAN-X antisera. From a total of 2320 spermatozoa examined by either conventional or confocal microscopy, 50.1% exhibited labeling with the SPAN-X antisera. Incomplete permeabilization of the nucleus and antibody inaccessibility caused by positional effects of the spermatozoa on slides cannot be excluded as an explanation for the low percentage of SPAN-X immunoreactive craters. However, crater staining throughout the sperm head was observed by z-sectioning using confocal microscopy. To our knowledge, SPAN-X is the only protein that shows a 50% segregation in ejaculated spermatozoa of any species.

The localization of SPAN-X to 50% of spermatozoa and its X-linked expression by haploid spermatids initially suggested that SPAN-X might be associated with only X-bearing spermatozoa. However, dual labeling of spermatozoa utilizing FISH for the X chromosome and indirect immunofluorescence for the SPAN-X protein demonstrated that SPAN-X was equally distributed between X- and Y-bearing spermatozoa, suggesting that SPAN-X and/or protein is shared within spermatid cohorts in the testis.

The mechanism by which SPAN-X is distributed to 50% of spermatozoa remains a fundamental question of great interest. One possibility is that all spermatids in the testis contain SPAN-X and, due to incomplete sperm maturation only half of the cells retain the protein at the time of ejaculation. Nuclear condensation and loss of residual cytoplasm occurs in spermatozoa as they mature in the epididymis (Bedford, J. M. (1973). Components of sperm maturation in the human epididymis. *Adv. Biosci.* 10, 145-155. Bedford, J. M., Bent, M. J., and Calvin, H. (1973). Variations in the structural character and stability of the nuclear chromatin in morphologically normal human spermatozoa. *J. Reprod. Fert.* 33, 19-29) and loss of SPAN-X in half of the cells may relate to this process. The possibility that the presence or absence of SPAN-X protein might affect the fertilizing ability of ejaculated spermatozoa remains an intriguing prospect. A second possibility is that SPAN-X could be expressed and transported within half of the spermatid cohorts in the testis. Experiments are in progress to further determine the incidence and function of SPAN-X during spermiogenesis and transport through the male and female reproductive

tracts and to follow the fate of SPAN-X at fertilization.

7. EXAMPLE: IDENTIFICATION OF SPAN-X HOMOLOGS IN DIFFERENT SPECIES

7.1. MATERIALS AND METHODS

Northern Analysis

Northern analysis was performed on Northern blots containing testicular mRNA from 4 species, human, cynomologus macaque, rat and mouse. A ^{32}P-labeled SPAN-X cDNA probe was used for hybridization. Following hybridization, the membrane was washed to a final stringency of 0.2X SSC and 0.2% SDS at 60°C. The membrane was exposed to autoradiographic film and the film processed to reveal hybridized bands.

7.2. RESULTS

7.2.1. MULTIPLE SPECIES NORTHERN ANALYSIS REVEALS A SPAN-X HOMOLOG IN RAT TESTIS

A broad band of approximately 0.6 kb was observed in human and rat testis when the membrane was probed with a SPAN-X cDNA probe. Very low signal was observed in the mouse and macaque. (FIG. 17)

8. EXAMPLE: LOCALIZATION OF SPAN-X IN HUMAN AND CHIMPANZEE TESTES

8.1. METHODS

Indirect immunofluorescence analysis was performed on human and chimpanzee testicular tissue to examine the localization of SPAN-X during spermatogenesis. Testes were obtained from human patients undergoing elective orchiectomies. Testes were sliced once with a razor blade and immersed in neutral buffered formalin (4%) solution (Sigma) for one hour. The tissue was minced, placed into fresh formalin overnight, dehydrated in a graded series of ethanols, cleared in xylene, and embedded in paraffin. 2.5 mm thick sections were cut and mounted onto slides. Chimpanzee testes sections mounted on slides were obtained from the Coulston Foundation in Alamogordo, New Mexico. Sections were depariffinized and rehydrated before immunostaining. To permeabilize the cells, the slides were incubated in methanol and washed with PBS. Non-specific protein binding sites were blocked by incubating the slides in PBS with 10% NDS. Slides were incubated with mouse

mAb MHS-10 (1:250) and guinea pig antiserum raised against recSPAN-X (1:250) diluted in PBS with 1% NDS (PBS-NS). The slides were washed and incubated with TRITC-conjugated $F(Ab)_2$ fragments of donkey anti-mouse IgG (1:200) and FITC-conjugated donkey anti-guinea pig IgG (1:200) (Jackson ImmunoResearch) in PBS-NDS. Slides were washed with PBS and mounted with Slow Fade Light (Molecular Probes, Eugene, Oregon) containing DAPI II counterstain (Vysis). Cells were observed by differential interference contrast (DIC) and epifluorescence microscopy using a Zeiss axiophot microscope. Individual blue (DAPI), green (FITC) and red (TRITC) fluorescent images were obtained using a digital camera (Hamamatsu) and compiled using Openlab software (Improvision Inc., Boston, MA).

8.2. RESULTS

In elongating and elongated spermatids, SPAN-X protein was observed as green fluorescence in the cytoplasmic droplet of elongating and elongated spermatids in both the human and chimpanzee testes. In round spermatids, occasional staining of the nuclear membrane was observed. No immunostaining was observed in other cell types of the seminiferous epithelium or the interstitium. The positive control protein SP-10, recognized by the mouse MHS-10 mAb, was observed as red fluorescent staining in the developing acrosome of round and elongating spermatids in both human and chimpanzee seminiferous tubules. The nuclear chromatin is stained blue with DAPI counterstain. (FIG. 16).

The invention is not to be limited in scope by the specific embodiments described which are intended as single illustrations of individual aspects of the invention, and functionally equivalent methods and components are within the scope of the invention. Indeed various modifications of the invention, in addition to those shown and described herein will become apparent to those skilled in the art from the foregoing description and accompanying drawings. Such modifications are intended to fall within the scope of the appended claims.

All references cited herein are incorporated herein by reference for all purposes.

WHAT IS CLAIMED IS:

1. An isolated nucleic acid molecule comprising a nucleotide sequence which encodes a protein comprising the amino acid sequence of SEQ ID NO:2 or SEQ ID NO:4.

2. An isolated nucleic acid molecule which comprises nucleotide sequence of SEQ ID NO: 1 or 3.

3. An isolated nucleic acid molecule derived from a mammalian genome that:

(a) hybridizes under highly stringent conditions to the nucleotide sequence of SEQ ID NO: 1 or 3; and

(b) encodes a gene product which is expressed in nuclear vacuoles of X-bearing and Y-bearing haploid sperm.

4. An isolated nucleic acid which comprises the nucleotide sequence of SEQ ID NO:20.

5. A nucleotide vector containing the nucleotide sequence of Claim 1, 2, or 3.

6. An expression vector containing the nucleotide sequence of Claim 1, 2, or 3 in operative association with a nucleotide regulatory sequence that controls expression of the nucleotide sequence in a host cell.

7. A genetically engineered host cell that contains the nucleotide of Claim 1, 2, or 3 in operative association with a nucleotide regulatory sequence that controls expression of the nucleotide sequence in the host cell.

8. An isolated SPAN-X protein having the amino acid sequence of SEQ ID NO: 2 or 4.

9. An isolated SPAN-X protein having a molecular weight (KD) and pI as follows: (i) MW 22.8, pI 4.99; (ii) MW 21.9, pI 5.02; (iii) MW 20.9, pI 5.00; (iv) protein of MW 23.2, pI 5.03; (v) MW 22.6, pI 5.06; (vi) MW 21.3, pI 5.09; (vii) MW 20.6, pI 5.06; (viii) MW 23.5, pI 5.09; (ix) MW 23.0, pI 6.30; (x) MW 22.4, pI 6.24; (xi) MW 22.1, pI 5.20; (xii) MW 22.8, pI 4.99; (xiii) MW 21.5, pI 5.24; (xiv) MW 21.6, pI 5.30; (xv) MW 20.1, pI 5.26; (xvi) MW 20.3, pI 6.23; (xvii) MW 20.2, pI 5.28; (xviii) MW 21.6, pI 5.54;

MW 21.3, pI 5.53; or (xix) MW 19.8, pI 5.44.

10. An antibody that immunospecifically binds the polypeptide of Claim 7.

11. A kit comprising an anti-SPAN-X antibody and a cryopreservative.

12. An isolated polynucleotide encoding a polypeptide comprising the amino acid sequence of SEQ ID NO: 17.

13. A polynucleotide comprising the polynucleotide of Claim 11 operatively associated with a nucleic acid encoding a polypeptide heterologous to SPAN-X.

14. An isolated polypeptide comprising the amino acid sequence of SEQ ID NO:17.

15. A fusion protein comprising the polypeptide of Claim 17 operatively associated with a heterologous polypeptide.

16. A method of importing a biologically active molecule into a cell *in vitro* or *ex vivo* comprising administering to the cell a complex comprising the molecule linked to an importation competent nuclear localization signal sequence, thereby importing the molecule into the nucleus of the cell.

17. The method of claim 16 wherein the nuclear localization signal sequence comprises the amino acid sequence set forth in SEQ ID NO:17, and wherein the presence of the nuclear localization signal sequence of SEQ ID NO:17 results in rapid transfer to nucleus.

18. The method of claim 16 wherein the nuclear localization signal sequence comprises the amino acid sequence set forth in SEQ ID NO:15, and wherein the presence of the nuclear localization signal sequence of SEQ ID NO:15 results in rapid transfer to nucleus.

19. The method of claim 16 wherein the nuclear localization signal sequence comprises the amino acid sequence set forth in SEQ ID NO:16, and wherein the presence of the nuclear localization signal sequence of SEQ ID NO:16 results in rapid transfer to

nucleus.

20. A method for separating SPAN-X-containing mammalian spermatozoa from SPAN-X- minus/lacking mammalian spermatozoa comprising:

(a) contacting mammalian spermatozoa with an antibody to SPAN-X; and

(b) purifying said SPAN-X-containing mammalian spermatozoa which bind to said antibody.

21. The method of claim 20, wherein the purified SPAN-X-containing mammalian spermatozoa may be used for fertilization including artificial insemination, intrauterine insemination or *in-vitro* fertilization.

22. A method for separating SPAN-X-containing mammalian spermatozoa from SPAN-X- minus/lacking mammalian spermatozoa comprising:

(a) contacting a solid phase surface containing anti-SPAN-X antibody with semen nuclei sample comprising Y-bearing sperm nuclei and X-bearing sperm nuclei, for a time period and under conditions sufficient to allow binding of the SPAN-X-containing X-bearing and Y-bearing sperm nuclei to the antibody on the solid phase surface;

(b) removing nuclei that are not bound to the solid phase surface;

(c) optionally, repeating steps (a) and (b) above, until substantially all SPAN-X-containing X-bearing and Y-bearing nuclei are bound by antibody; and

(d) recovering the substantially pure preparation of SPAN-X-containing X-bearing and Y-bearing sperm nuclei not bound to the surface.

23. A method of gene therapy for a patient in need thereof comprising administering a therapeutically effective amount of a vector containing a gene of interest under control of the *span-x* promoter wherein said *span-x* promoter sequences are used to drive spermatid-specific expression of drugs or toxins.

24. A method of contraception for a patient in need thereof comprising administering a therapeutically effective amount of a vector containing a gene of interest under control of the *span-x* promoter, wherein said *span-x* promoter is used for contraception to drive spermatid expression of a toxin in the testis.

25. A method of gene therapy for a patient in need thereof comprising administering a therapeutically effective amount of a vector containing a gene of interest under control of the *span-x* promoter wherein said *span-x* promoter sequences are used to drive spermatid-specific expression of drugs or toxins.

26. A method of contraception for a patient in need thereof comprising administering a therapeutically effective amount of a vector containing a gene of interest under control of the *span-x* promoter, wherein said *span-x* promoter is used for contraception to drive spermatid expression of a toxin in the testis.

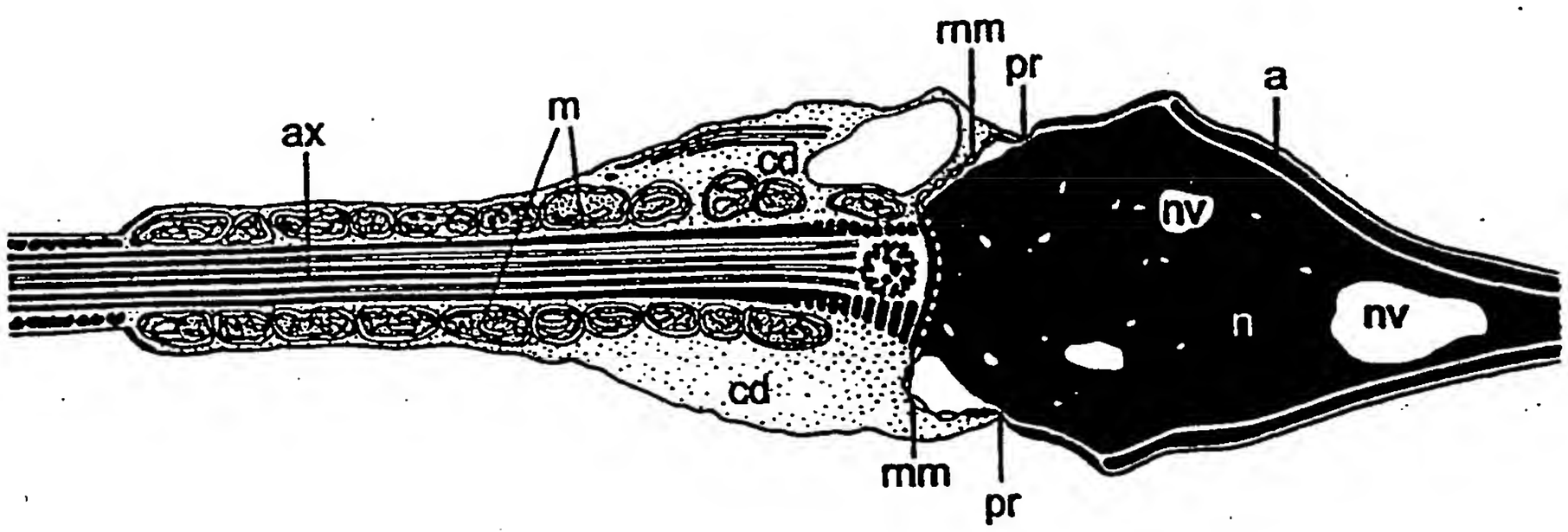


FIG. 1

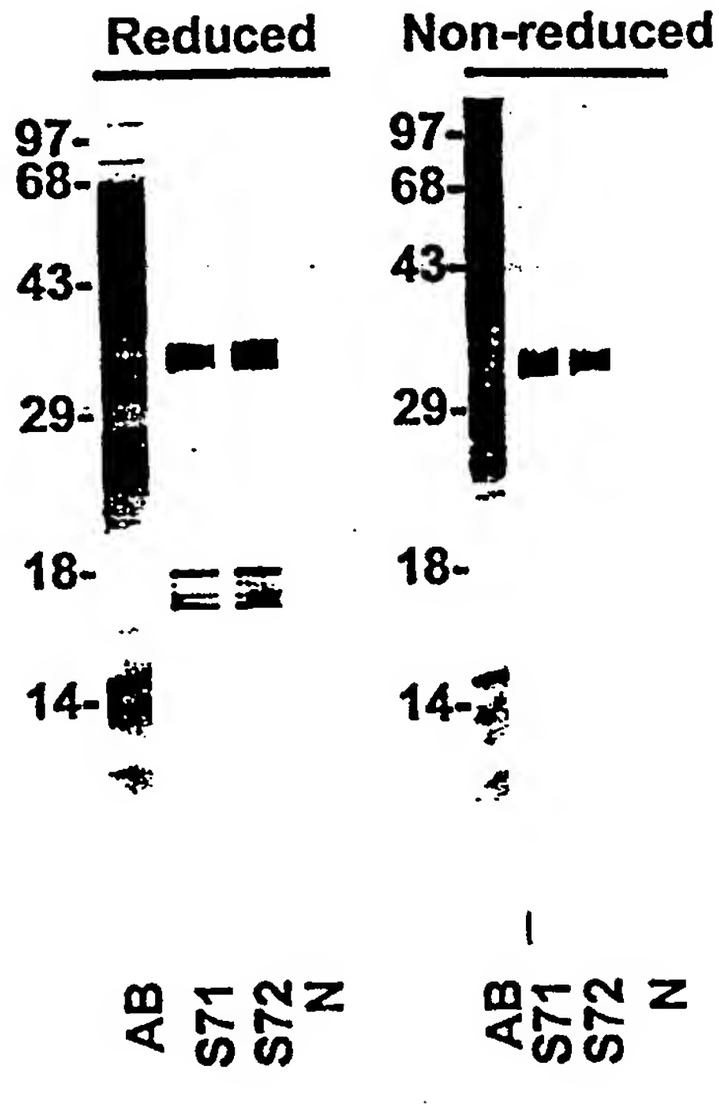


FIG. 2

A

```
          1                                                    50
SPAN-Xa   ---------- ---------- ---------- ---------- ----------
SPAN-Xb   GTCACCAGGA GGGTATGCAT AGGGAGGGCA AGAGCTCTGG GCCACTGCGA

          51                                                   100
SPAN-Xa   ---------- --------AAG CCTGCCACTG ACATTGAAGA ACCAATATAT
SPAN-Xb   AGATTCAAAA GCTCCAAAAA CCTACTGTAG ACATCGAAGA ACCAATATAT

          101                                                  150
SPAN-Xa   ACAATGGACA AACAATCCAG TGCCGGCGGG GTGAAGAGGA GCGTCCCCTG
SPAN-Xb   ACAATGGGCC AACAATCCAG TGTCCGCAGG CTGAAGAGGA GCGTCCCCTG

          151                                                  200
SPAN-Xa   TGATTCCAAC GAGGCCAACG AGATG..... .......... ...ATGCCGG
SPAN-Xb   TGAATCCAAC GAGGCCAACG AGGCCAATGA GGCCAACAAG ACGATGCCGG

          201                                                  250
SPAN-Xa   AGACCCCAAC TGGGGACTCA GACCCGCAAC CTGCTCCTAA AAAAATGAAA
SPAN-Xb   AGACCCCAAC TGGGGACTCA GACCCGCAAC CTGCTCCTAA AAAAATGAAA

          251                                                  300
SPAN-Xa   ACATCTGAGT CCTCGACCAT ACTAGTGGTT CGCTACAGGA GGAACTTTAA
SPAN-Xb   ACATCTGAGT CCTCGACCAT ACTAGTGGTT CGCTACAGGA GGAACGTGAA

          301                                                  350
SPAN-Xa   AAGAACATCT CCAGAGGAAC TGCTGAATGA CCACGCCCGA GAGAACAGAA
SPAN-Xb   AAGAACATCT CCAGAGGAAC TGGTGAATGA CCACGCCCGA GAGAACAGAA

          351                                                  400
SPAN-Xa   TCAACCCCCT CCAAATGGAG GAGGAGGAAT TCATGGAAAT AATGGTTGAA
SPAN-Xb   TCAACCCCGA CCAAATGGAG GAGGAGGAAT TCATAGAAAT AACGACTGAA

          401                                                  450
SPAN-Xa   ATACCTGCAA AGTAGCAAGA AGCTACATCT CTCAACCTTG GCAATGAAA
SPAN-Xb   AGACCTAAAA AGTAGCAAGA AGCTACATCC CTCAAACTTC GGCAATGAAA

          451                469
SPAN-Xa   ATAAAGTTTG AGAAGCTGA
SPAN-Xb   ATAAAGTTTG AGAAGCTGA
```

B

```
  1 MDKQSSAGGVKRSVPCDSNEANEM.....MPETPTGDSDPQPAPKKMKT
    | .|||   .||||||:||||||       ||||||||||||||||||
  1 MGQQSSVRRLKRSVPCESNEANEANEANKTMPETPTGDSDPQPAPKKMKT

 45 SESSTILVVRYRRNFKRTSPEELLNDHARENRINPLQMEEEEEFMEIMVEI
    |||||||||||||| |||||||.|||||||||| |||||||.||  |
 51 SESSTILVVRYRRNVKRTSPEELVNDHARENRINPDQMEEEEEFIEITTER

 95 PAK 97
    | |
101 PKK 103
```

FIG. 3

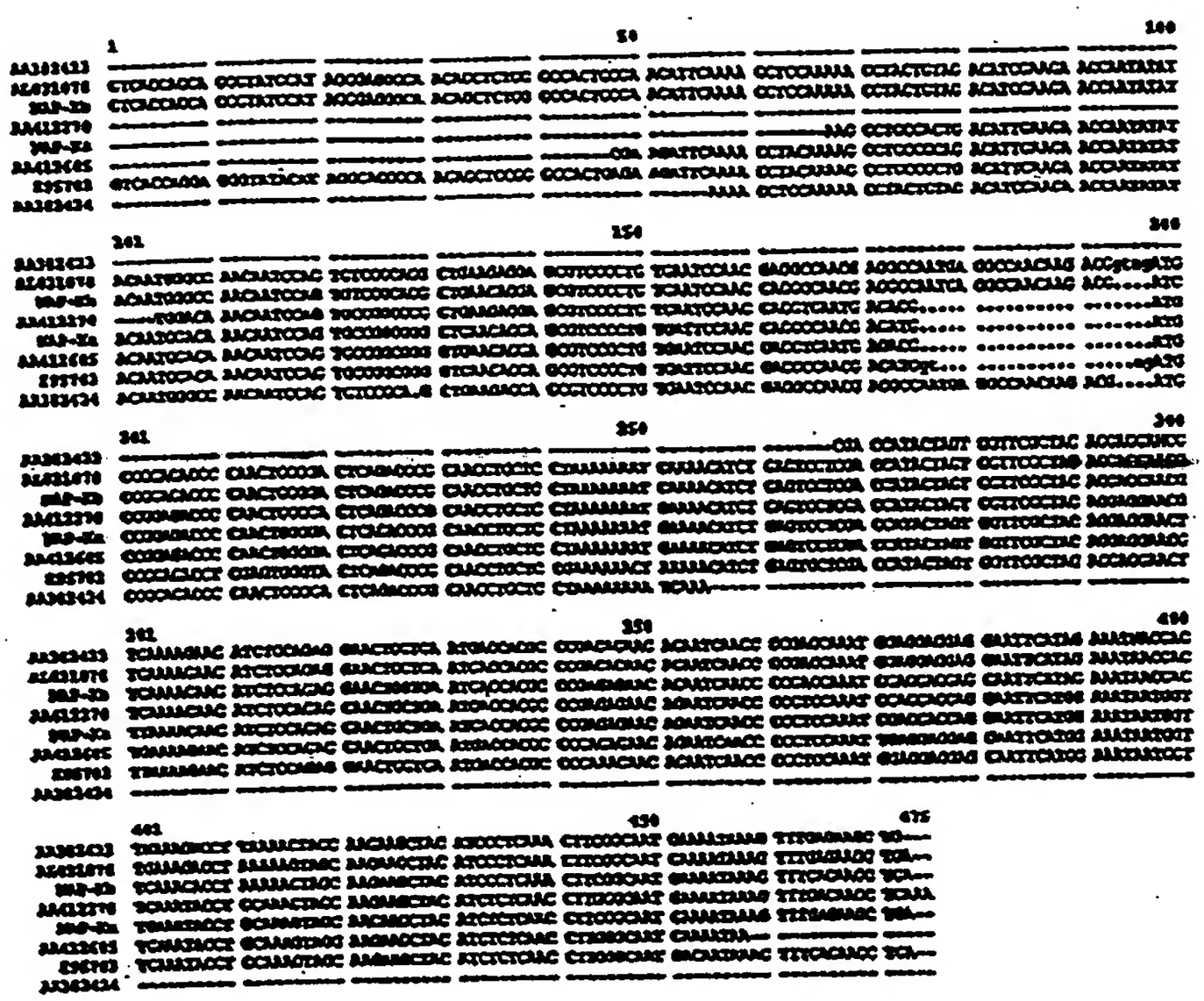


FIG. 4A

```
                1                                                    50
ai143898        ---------- ---------- ---------- ---------- ----------
ai962751        ---------- ---------- ---------- ---------- ----------
ai808260        ---------- ---------- ---------- ---------- ----------
 SPAN-Xb        GTCACCAGGA GGGTATGCAT AGGGAGGGCA AGAGCTCTGG GCCACTGCGA
aa382423        ---------- ---------- ---------- ---------- ----------
 SPAN-Xa        ---------- ---------- ---------- ---------- ----------
aa412270        ---------- ---------- ---------- ---------- ----------
ai208372        ---------- ---------- ---------- ---------- ----------
aa412605        ---------- ---------- ---------- ---------- --------CGA
aa382424        ---------- ---------- ---------- ---------- ----------

               51                                                   100
ai143898        ---------- ---------- ---------- ---------- ----------
ai962751        ---------- ---------- --------GGC ACGAGGAGCC TGCCACTGAC
ai808260        ---------- ---------- ---------- ---------- ----------
 SPAN-Xb        AGATTCAAAA GCTCCAAAAA CCTACTGTAG ACATCGAAGA ACCAATATAT
aa382423        ---------- ---------- ---------- ---------- ----------
 SPAN-Xa        ---------- --------AAG CCTGCCACTG ACATTGAAGA ACCAATATAT
aa412270        ---------- ---------- ---------- ---------- ----------
ai208372        ---------- ---------- ---------- ---------- ----------
aa412605        AGATTCAAAA CCTACAAAAG CCTGCCGCAG ACATTGAAGA ACCAATATAT
aa382424        ------AAAA GCTCCAAAAA CCTACTGTAG ACATCGAAGA ACCAATATAT

              101                                                   150
ai143898        ---------- ---------- ---------- ---------- ----------
ai962751        ATTGAAGAAC CAATATATAC AATGGACAAA CAATCCAGTG CCGGCGGGGT
ai808260        ---------- ---------- ---------- ---------- ----------
 SPAN-Xb        ACAATGGGCC AACAATCCAG TGTCCGCAGG CTGAAGAGGA GCGTCCCCTG
aa382423        ---------- ---------- ---------- ---------- ----------
 SPAN-Xa        ACAATGGACA AACAATCCAG TGCCGGCGGG GTGAAGAGGA GCGTCCCCTG
aa412270        ----TGGACA AACAATCCAG TGCCGGCGGG GTGAAGAGGA GCGTCCCCTG
ai208372        ---------- ---------- ------CGGG GTGAAGAGGA GCGTCCCCTG
aa412605        ACAATGGACA AACAATCCAG TGCCGGCGGG GTGAAGAGGA GCGTCCCCTG
aa382424        ACAATGGGCC AACAATCCAG TGTCCGCA.G CTGAAGAGGA GCGTCCCCTG

              151                                                   200
ai143898        ---------- ---------- ---------- ---------- ----------
ai962751        GAAGAGGAGC GTCCCCTGTG ATTCCAACGA GGCCAACGAG ATGATGCCGG
ai808260        ---------- -----TCCTG GCCTCTCCCT GTTTTCTTAA CAGATGCCGG
 SPAN-Xb        TGAATCCAAC GAGGCCAACG AGGCCAATGA GGCCAACAAG ACGATGCCGG
aa382423        ---------- ---------- ---------- ---------- ----------
 SPAN-Xa        TGATTCCAAC GAGGCCAACG AGATG..... .......... ...ATGCCGG
aa412270        TGAATCCAAC GAGGTGAATG AGACG..... .......... ...ATGCCGG
ai208372        TGATTCCAAC GAGGCCAACG AGATG..... .......... ...ATGCCGG
aa412605        TGAATCCAAC GAGGTGAATG AGACG..... .......... ...ATGCCGG
aa382424        TGAATCCAAC GAGGCCAACG AGGCCAATGA GGCCAACAAG ACGATGCCGG

              201                                                   250
ai143898        ---------- ---------- ---------- --GCTCCGAA AAAACTAAAA
ai962751        AGACCCCAAC TGGGGACTCA GACCCGCAAC CTGCTCCTAA AAAAATGAAA
ai808260        AGACCTCGAG TGGGTACTCA GACCCGCAAC CTGCTCCGAA AAAACTAAAA
 SPAN-Xb        AGACCCCAAC TGGGGACTCA GACCCGCAAC CTGCTCCTAA AAAAATGAAA
aa382423        ---------- ---------- ---------- ---------- ----------
 SPAN-Xa        AGACCCCAAC TGGGGACTCA GACCCGCAAC CTGCTCCTAA AAAAATGAAA
aa412270        AGACCCCAAC TGGGGACTCA GACCCGCAAC CTGCTCCTAA AAAAATGAAA
ai208372        AGACCCCAAC TGGGGACTCA GACCCGCAAC CTGCTCCTAA AAAAATGAAA
aa412605        AGACCCCAAC TGGGGACTCA GACCCGCAAC CTGCTCCTAA AAAAATGAAA
aa382424        AGACCCCAAC TGGGGACTCA GACCCGCAAC CTGCTCCTAA AAAAAATGAA
```

FIG. 4B

```
                251                                                       300
ai143898    ACATTTGAGT CCTCGACCAT ACTAGTGGTT CGCTACAGGA GGAACTTTAA
ai962751    ACATCTGAGT CCTCGACCAT ACTAGTGGTT CGCTACAGGA GGAACTTTAA
ai808260    ACATCTGAGT CCTCGACCAT ACTAGTGGTT CGCTACAGGA GGAACTTTAA
 SPAN-Xb    ACATCTGAGT CCTCGACCAT ACTAGTGGTT CGCTACAGGA GGAACGTGAA
aa382423    ---------- ---CGACCAT ACTAGTGGTT CGCTACAGGA GGANCGTGAA
 SPAN-Xa    ACATCTGAGT CCTCGACCAT ACTAGTGGTT CGCTACAGGA GGAACTTTAA
aa412270    ACATCTGAGT CCTCGACCAT ACTAGTGGTT CGCTACAGGA GGAACGTGAA
ai208372    ACATCTGAGT CCTCGACCAT ACTAGTGGTT CGCTACAGGA GGAACTTTAA
aa412605    ACATCTGAGT CCTCGACCAT ACTAGTGGTT CGCTACAGGA GGAACGTGAA
aa382424    A--------- ---------- ---------- ---------- ----------

                301                                                       350
ai143898    AAGAACATCT CCAGAGGAAC TGGTGAATGA CCCCGCCCGA AAGAACAGAA
ai962751    AAGAACATCT CCAGAGGAAC TGCTGAATGA CCACGCCCGA GAGAACAGAA
ai808260    AAGAACATCT CCAGAGGAAC TGGTGAATGA CCACGCCCGA AAGAACAGAA
 SPAN-Xb    AAGAACATCT CCAGAGGAAC TGGTGAATGA CCACGCCCGA GAGAACAGAA
aa382423    AAGAACATCT CCAGAGGAAC TGGTGAATGA CCACGCCCGA GAGAACAGAA
 SPAN-Xa    AAGAACATCT CCAGAGGAAC TGCTGAATGA CCACGCCCGA GAGAACAGAA
aa412270    AAGAACATCT CCAGAGGAAC TGCTGAATGA CCACGCCCGA GAGAACAGAA
ai208372    AAGAACATCT CCAGAGGAAC TGCTGAATGA CCACGCCCGA GAGAACAGAA
aa412605    AAGAACATCT CCAGAGGAAC TGCTGAATGA CCACGCCCGA GAGAACAGAA
aa382424    ---------- ---------- ---------- ---------- ----------

                351                                                       400
ai143898    TCAACCCCCT CCAAATGGAG GAGGAGGAAT TCATGGAAAT AATGGTTGAA
ai962751    TCAACCCCCT CCAAATGGAG GAGGAGGAAT TCATGGAAAT AATGGTTGAA
ai808260    TCAACCCCCT CCAAATGGAG GAGGAGGAAT TCATGGAAAT AATGGTTGAA
 SPAN-Xb    TCAACCCCGA CCAAATGGAG GAGGAGGAAT TCATAGAAAT AACGACTGAA
aa382423    TCAACCCCGA CCAAATGGAG GAGGAGGAAT TCATAGAAAT NACGACTNAA
 SPAN-Xa    TCAACCCCCT CCAAATGGAG GAGGAGGAAT TCATGGAAAT AATGGTTGAA
aa412270    TCAACCCCCT CCAAATGGAG GAGGAGGAAT TCATGGAAAT AATGGTTGAA
ai208372    TCAACCCCCT CCAAATGGAG GAGGAGGAAT TCATGGAAAT AATGGTTGAA
aa412605    TCAACCCCCT CCAAATTGAG GAGGAGGAAT TCATGGAAAT AATGGTTGAA
aa382424    ---------- ---------- ---------- ---------- ----------

                401                                                       450
ai143898    ATACCTGCAA AGTAGCAAGA AGGTACATCT CTCAACCTTG GGCAATGCCA
ai962751    ATACCTGCAA AGTAGCAAGA AGCTACATCT CTCAACCTTG GGCAATGAAA
ai808260    ATACCTGCAA AGTAGCAAGA AGCTACATCT CTCAACCTTG GGCAACCCCA
 SPAN-Xb    AGACCTAAAA AGTAGCAAGA AGCTACATCC CTCAAACTTC GGCAATGAAA
aa382423    AGNCCTTAAA AGTAGCAAGA AGCTACATCC CTCAAACTTC GGCAATGAAA
 SPAN-Xa    ATACCTGCAA AGTAGCAAGA AGCTACATCT CTCAACCTTG GGCAATGAAA
aa412270    ATACCTGCAA AGTAGCAAGA AGCTACATCT CTCAACCTTG GGCAATGAAA
ai208372    ATACCTGCAA AGTAGCAAGA AGCTACATCT CTCAACCTTG GGCAATGAAA
aa412605    ATACCTGCAA AGTAGCAAGA AGCTACATCT CTCAACCTTG GGCAATGAAA
aa382424    ---------- ---------- ---------- ---------- ----------

                451                                                  493
ai143898    ATAAAGTTTG AGAAGCTGAA AAAAAAAAAA AAAAAAAAAA AAA
ai962751    ATAAAGTTTG AGAAGCTGAA AAAAAAAAAA ---------- ---
ai808260    ATAAAGTTTG AGAAGCTGAT GGCTGTGTAA A--------- ---
 SPAN-Xb    ATAAAGTTTG AGAAGCTGA- ---------- ---------- ---
aa382423    ATAAAGTTTG AGAAGCTG-- ---------- ---------- ---
 SPAN-Xa    ATAAAGTTTG AGAAGCTGA- ---------- ---------- ---
aa412270    ATAAAGTTTG AGAAGCTGAA A--------- ---------- ---
ai208372    ATAAAGTTTG AGAAGCTGAT GGCTGTGTAA A--------- ---
aa412605    ATAA------- ---------- ---------- ---------- ---
aa382424    ---------- ---------- ---------- ---------- ---
```

FIG. 4B (CONTINUED)

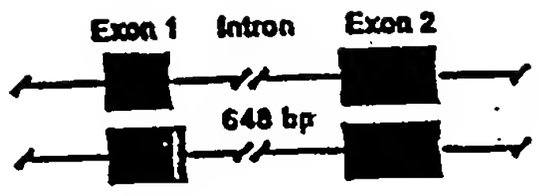


FIG. 4C

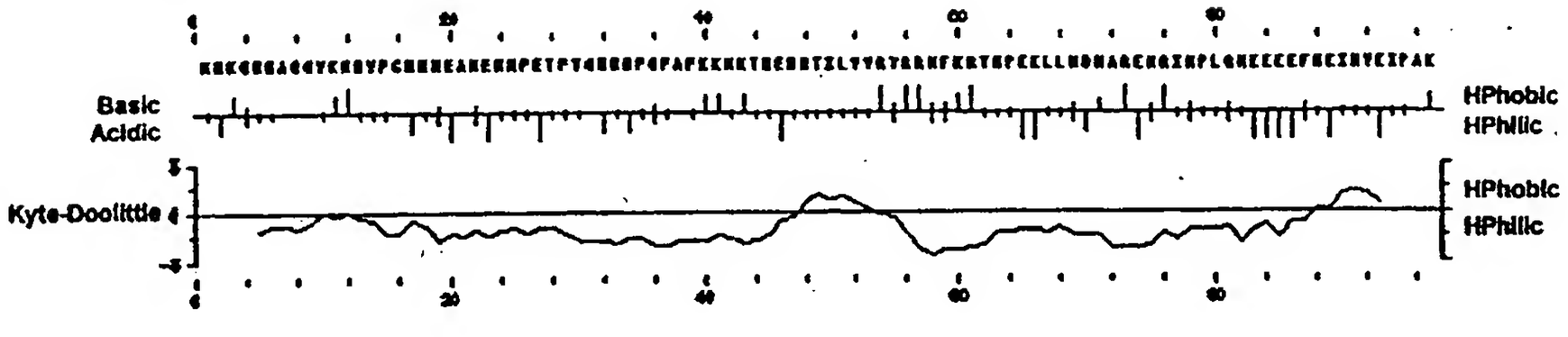

SPAN-Xa
Basic
Acidic
HPhobic
HPhilic
Kyte-Doolittle
HPhobic
HPhilic
SPAN-Xb
Basic
Acidic
HPhobic
HPhilic
Kyte-Doolittle
HPhobic
HPhilic

FIG. 5

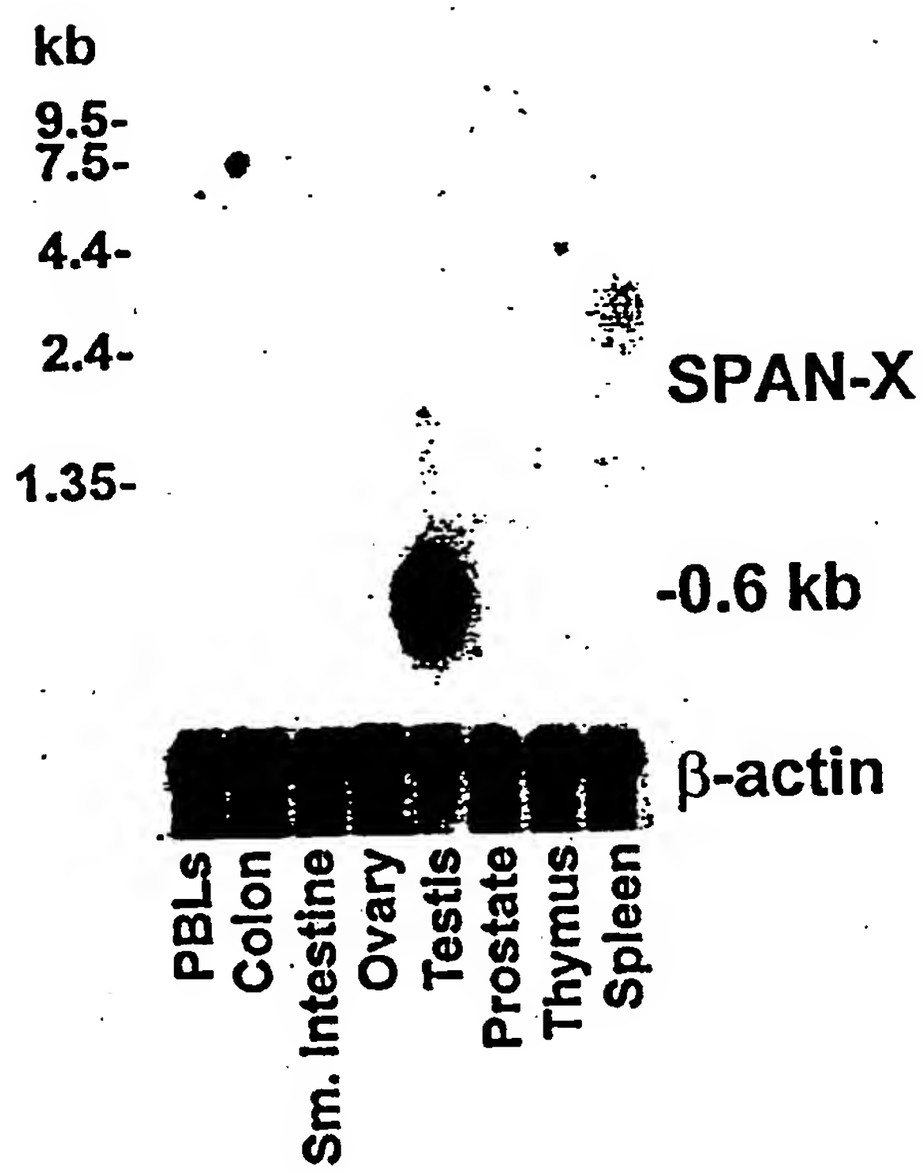

kb
9.5-
7.5-
4.4-
2.4-
1.35-
SPAN-X
-0.6 kb
β-actin
PBLs
Colon
Sm. Intestine
Ovary
Testis
Prostate
Thymus
Spleen

FIG. 6

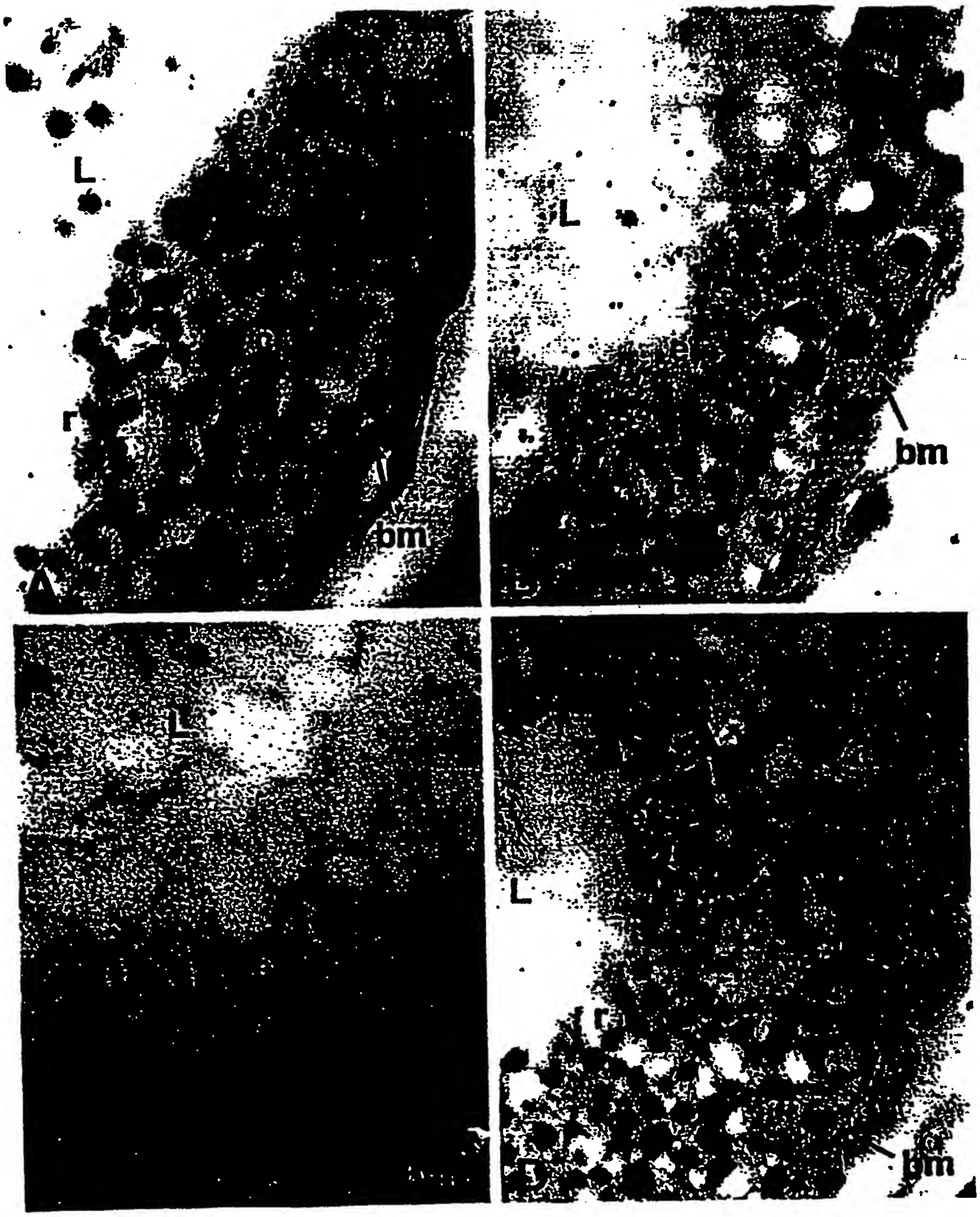


FIG. 7

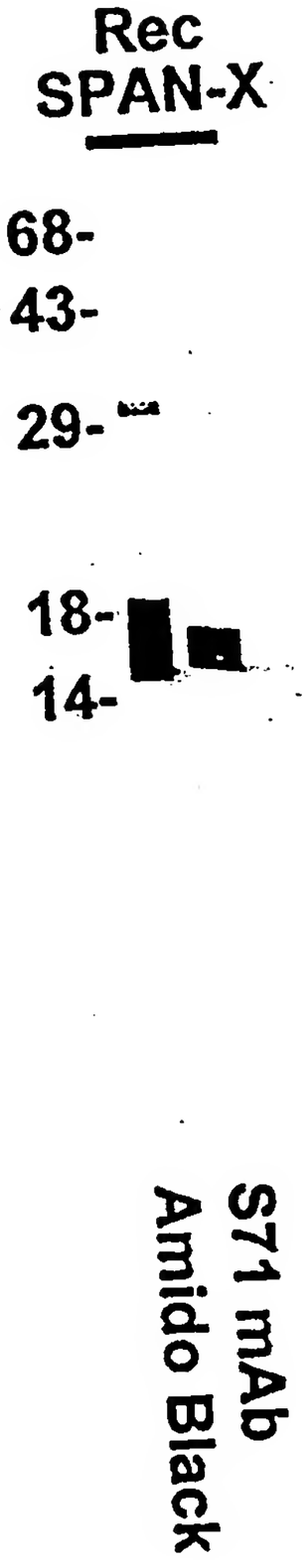


FIG. 8

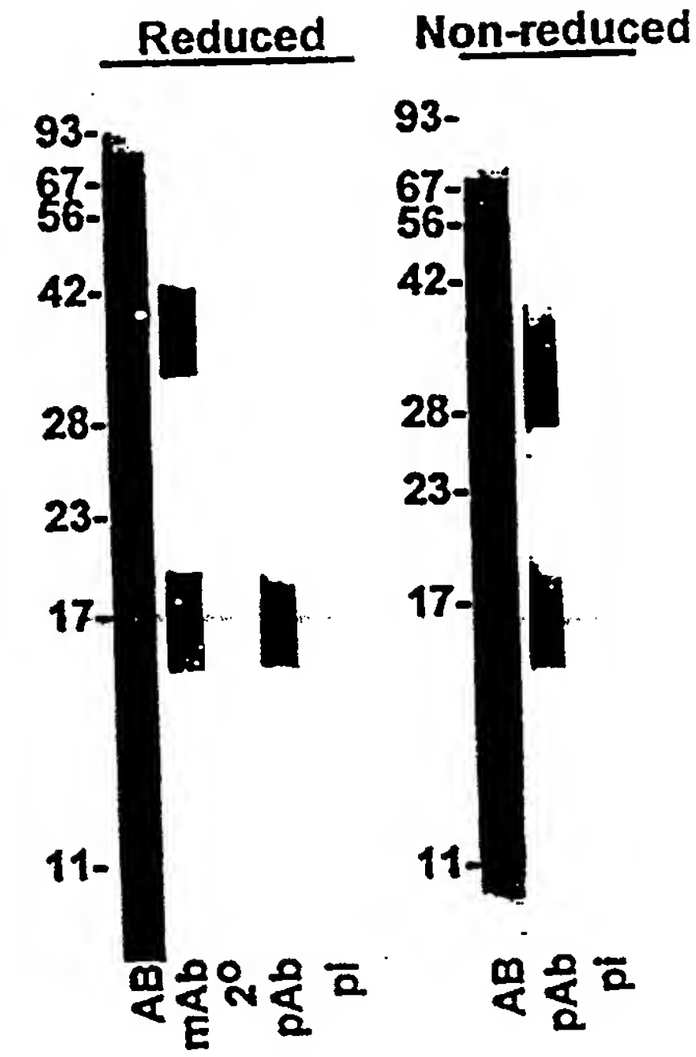

Reduced
Non-reduced
93-
67-
56-
42-
28-
23-
17-
11-
AB
mAb
2°
pAb
pi
AB
pAb
pi

FIG. 9

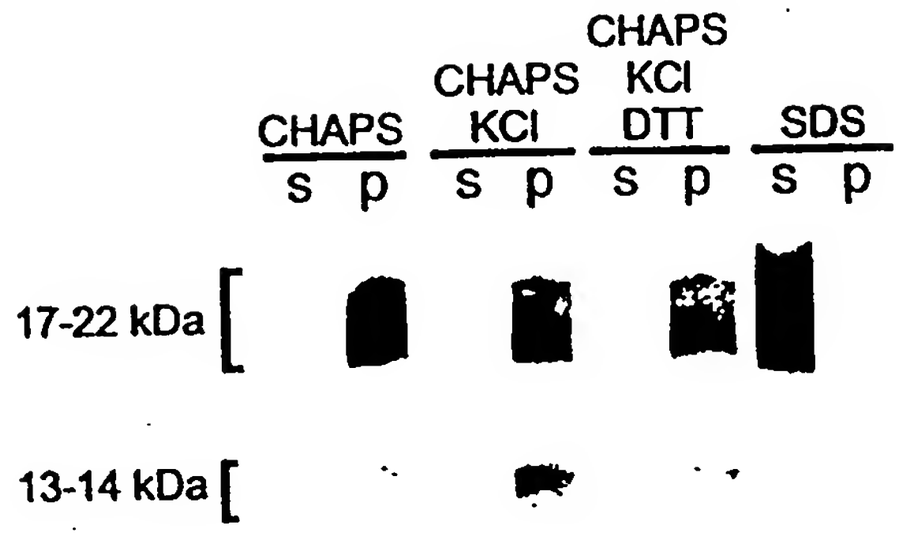


FIG. 10

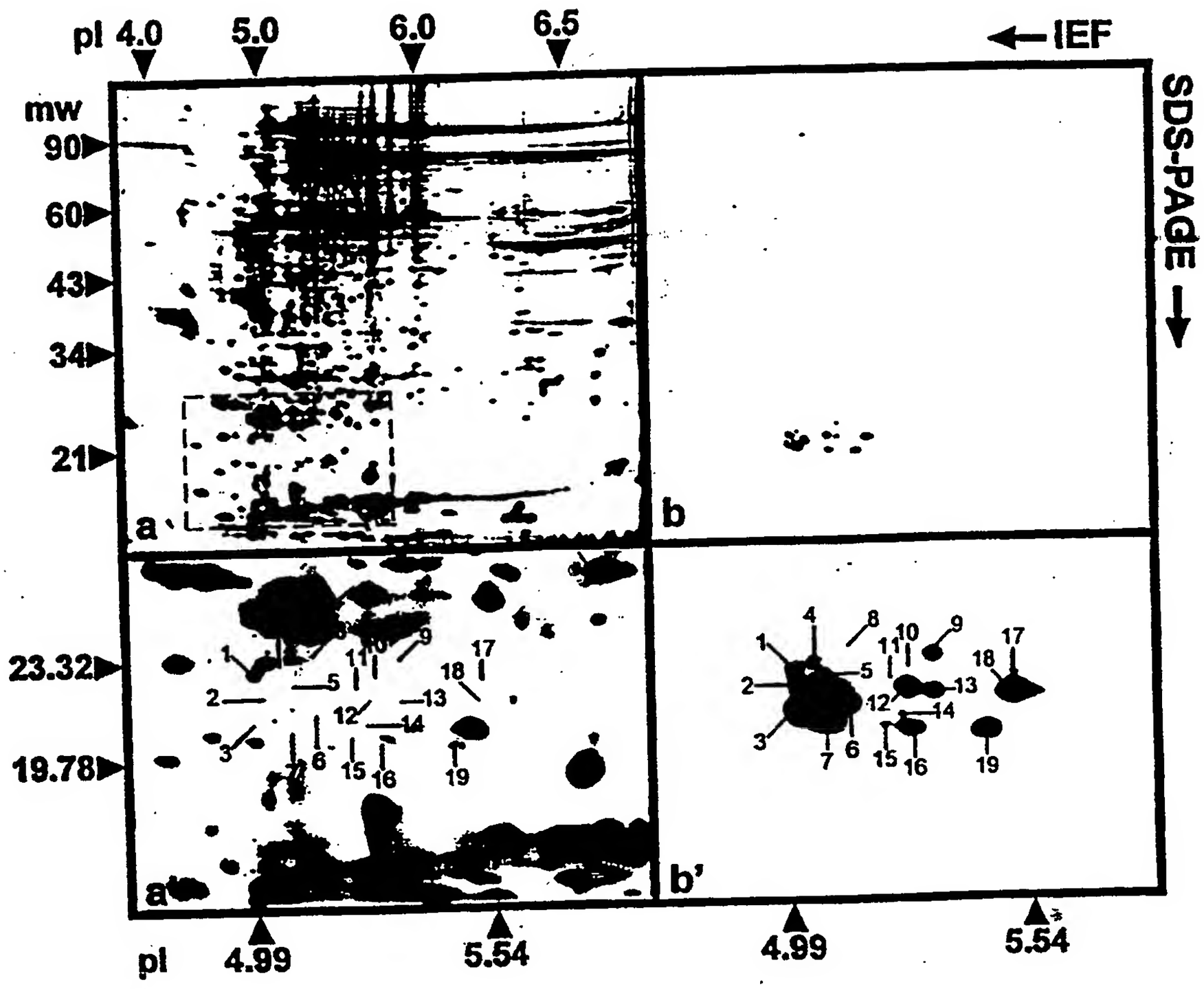

pI 4.0
5.0
6.0
6.5
IEF
SDS-PAGE
mw
90
60
43
34
21
23.32
19.78
a
b
a'
b'
pI
4.99
5.54
4.99
5.54

FIG. 11

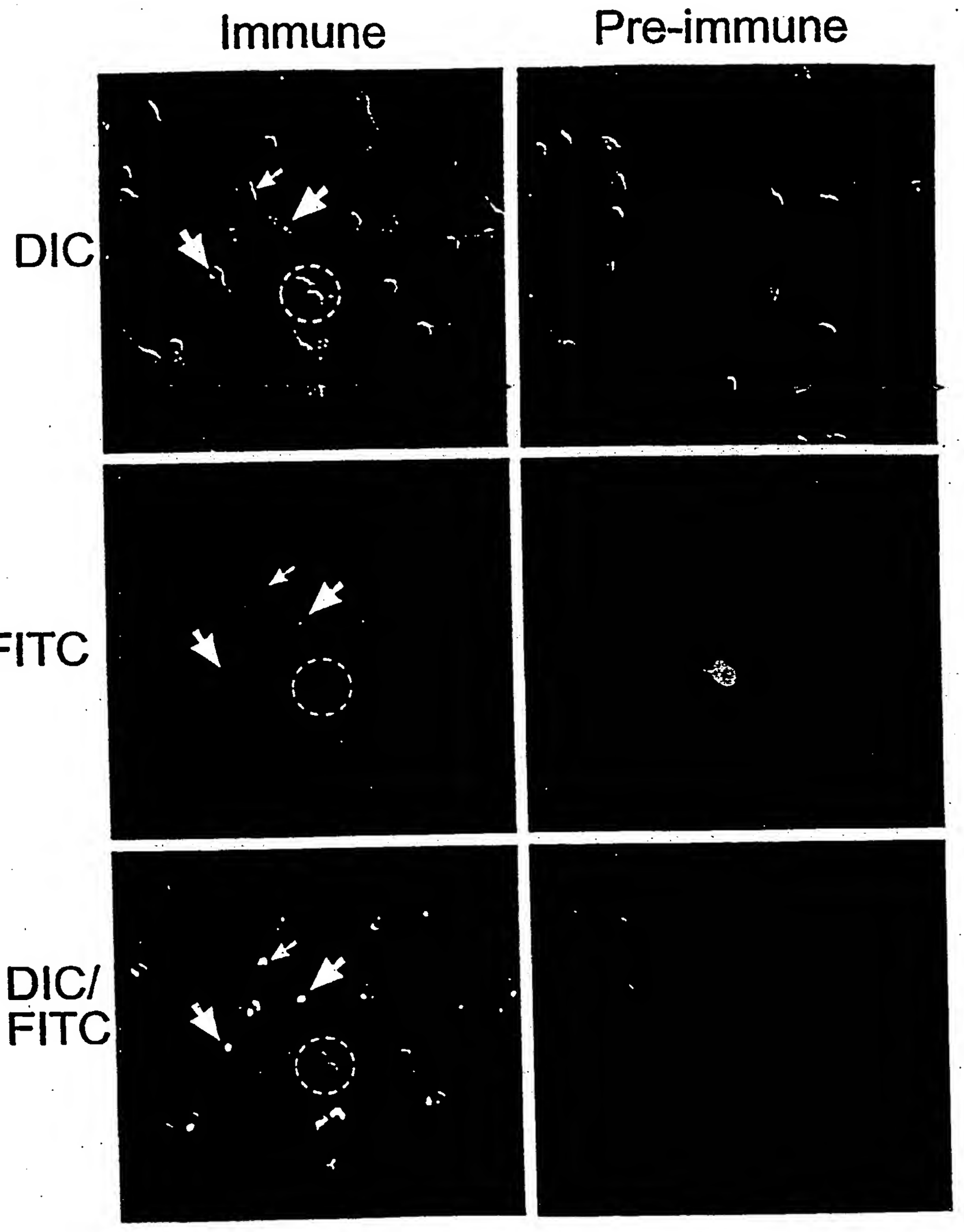


FIG. 12

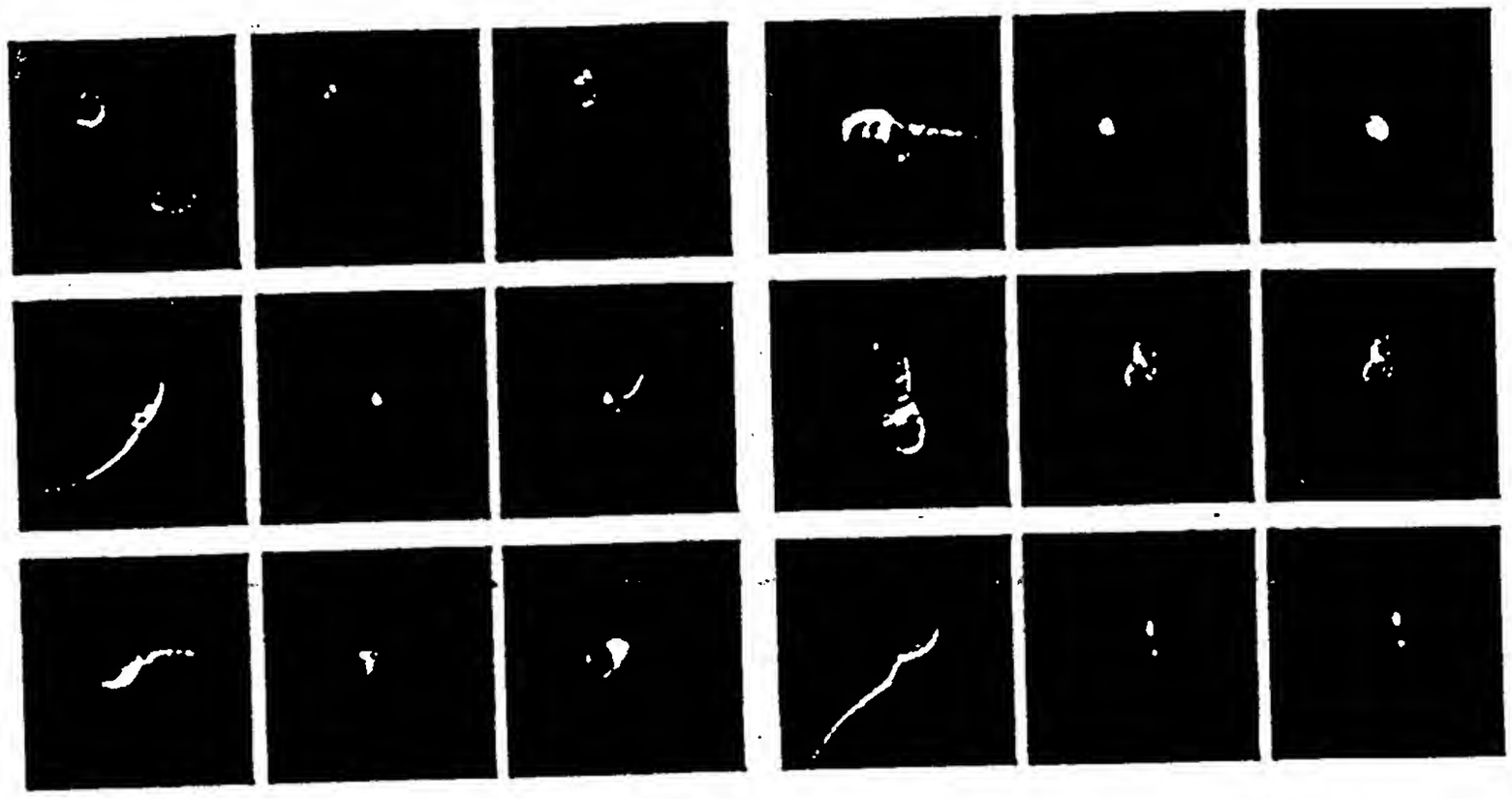


FIG. 13

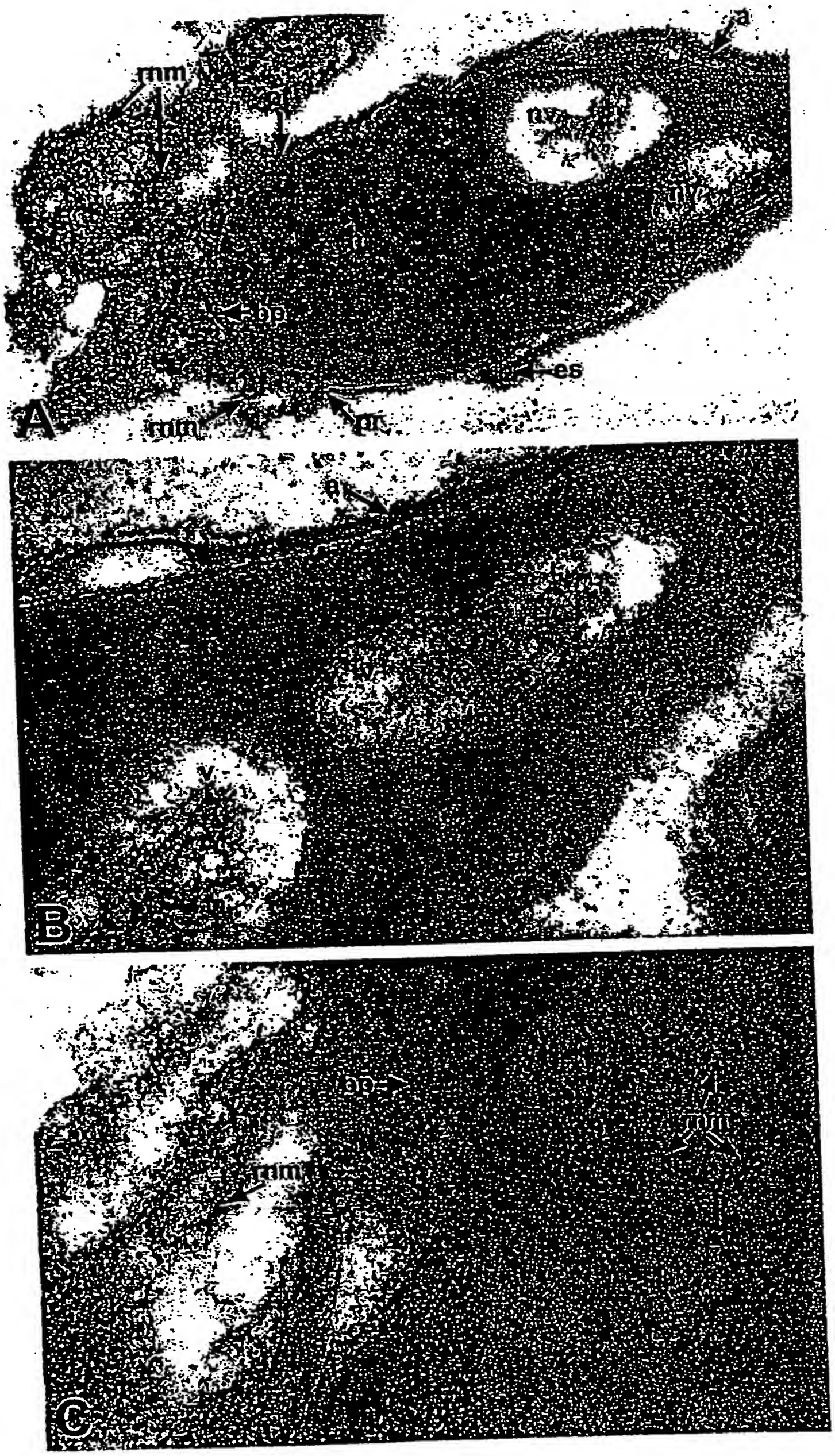


FIG. 14

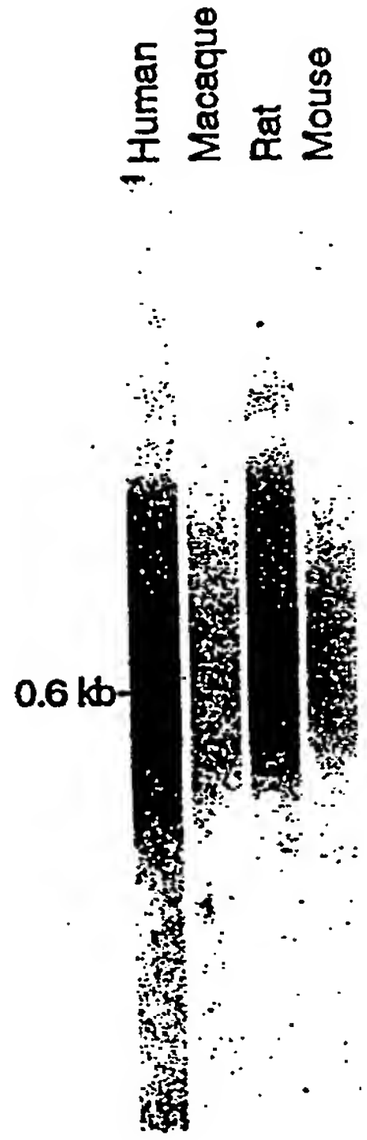


FIG. 15

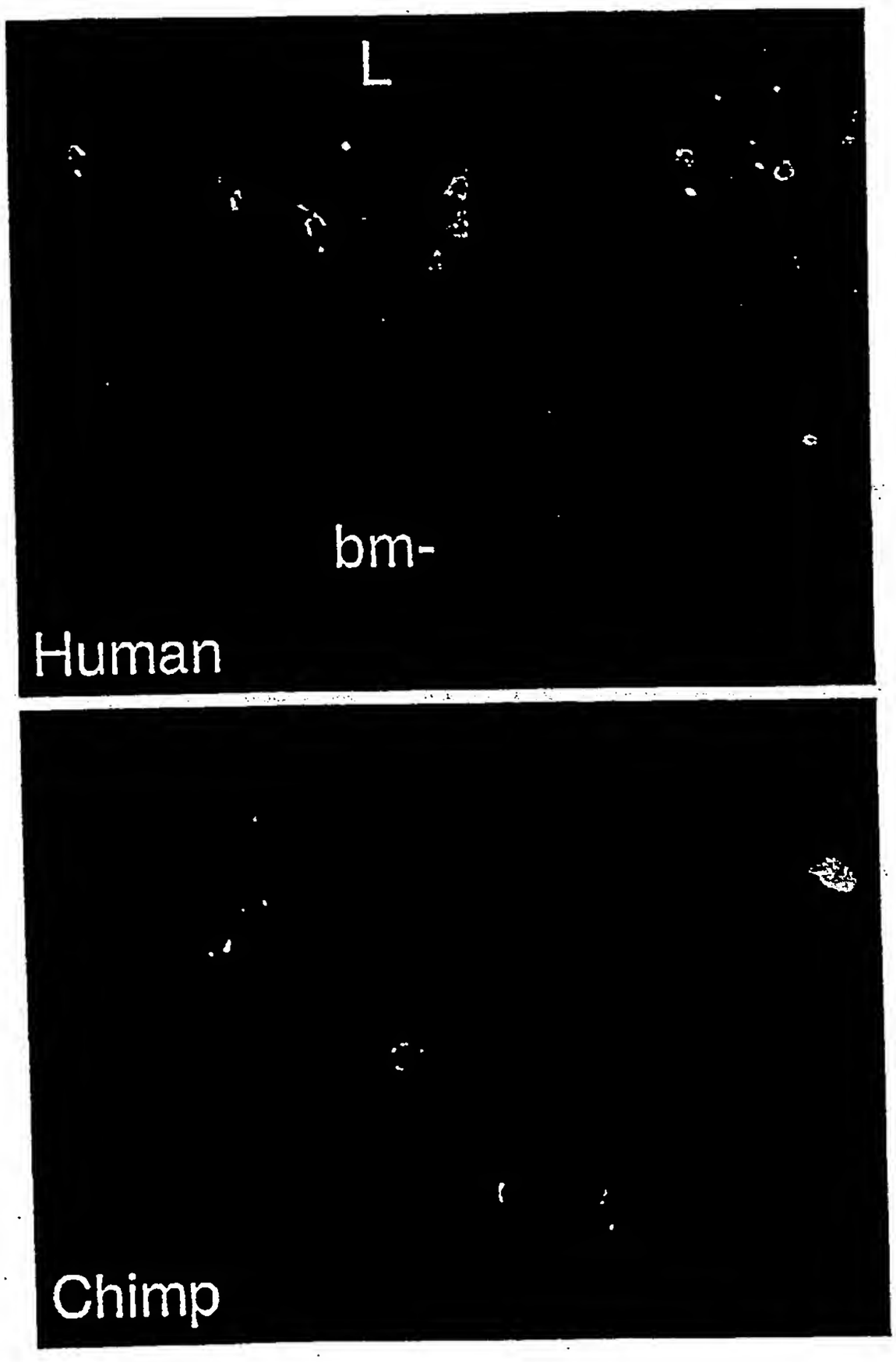


FIG. 16

SEQUENCE LISTING

<110> University of Virginia Patent Foundation

<120> NUCLEIC ACIDS AND PROTEINS OF THE HUMAN SPAN-X GENE AND USES THEREOF

<130> 9426-009-228

<140> Not Assigned
<141> 1998-10-22

<150> 60/105,324
<151> 1998-10-23

<160> 27

<170> FastSEQ for Windows Version 4.0

<210> 1
<211> 384
<212> DNA
<213> Homo Sapiens

<220>
<221> CDS
<222> (37)...(330)

<400> 1

```
aagcctgcca ctgacattga agaaccaata tataca atg gac aaa caa tcc agt        54
                                          Met Asp Lys Gln Ser Ser
                                           1               5

gcc ggc ggg gtg aag agg agc gtc ccc tgt gat tcc aac gag gcc aac       102
Ala Gly Gly Val Lys Arg Ser Val Pro Cys Asp Ser Asn Glu Ala Asn
            10                  15                  20

gag atg atg ccg gag acc cca act ggg gac tca gac ccg caa cct gct       150
Glu Met Met Pro Glu Thr Pro Thr Gly Asp Ser Asp Pro Gln Pro Ala
        25                  30                  35

cct aaa aaa atg aaa aca tct gag tcc tcg acc ata cta gtg gtt cgc       198
Pro Lys Lys Met Lys Thr Ser Glu Ser Ser Thr Ile Leu Val Val Arg
    40                  45                  50

tac agg agg aac ttt aaa aga aca tct cca gag gaa ctg ctg aat gac       246
Tyr Arg Arg Asn Phe Lys Arg Thr Ser Pro Glu Glu Leu Leu Asn Asp
55                  60                  65                  70

cac gcc cga gag aac aga atc aac ccc ctc caa atg gag gag gag gaa       294
His Ala Arg Glu Asn Arg Ile Asn Pro Leu Gln Met Glu Glu Glu Glu
                75                  80                  85

ttc atg gaa ata atg gtt gaa ata cct gca aag tag caagaagcta            340
Phe Met Glu Ile Met Val Glu Ile Pro Ala Lys  *
            90                  95
```

catctctcaa ccttgggcaa tgaaaataaa gtttgagaag ctga 384

<210> 2
<211> 97
<212> PRT
<213> Homo Sapiens

<400> 2
Met Asp Lys Gln Ser Ser Ala Gly Gly Val Lys Arg Ser Val Pro Cys
1 5 10 15
Asp Ser Asn Glu Ala Asn Glu Met Met Pro Glu Thr Pro Thr Gly Asp
20 25 30
Ser Asp Pro Gln Pro Ala Pro Lys Lys Met Lys Thr Ser Glu Ser Ser
35 40 45
Thr Ile Leu Val Val Arg Tyr Arg Arg Asn Phe Lys Arg Thr Ser Pro
50 55 60
Glu Glu Leu Leu Asn Asp His Ala Arg Glu Asn Arg Ile Asn Pro Leu
65 70 75 80
Gln Met Glu Glu Glu Glu Phe Met Glu Ile Met Val Glu Ile Pro Ala
85 90 95
Lys

<210> 3
<211> 469
<212> DNA
<213> Homo Sapiens

<220>
<221> CDS
<222> (104)...(415)

<400> 3
gtcaccagga gggtatgcat agggagggca agagctctgg gccactgcga agattcaaaa 60
gctccaaaaa cctactgtag acatcgaaga accaatatat aca atg ggc caa caa 115
Met Gly Gln Gln
1

tcc agt gtc cgc agg ctg aag agg agc gtc ccc tgt gaa tcc aac gag 163
Ser Ser Val Arg Arg Leu Lys Arg Ser Val Pro Cys Glu Ser Asn Glu
5 10 15 20

gcc aac gag gcc aat gag gcc aac aag acg atg ccg gag acc cca act 211
Ala Asn Glu Ala Asn Glu Ala Asn Lys Thr Met Pro Glu Thr Pro Thr
25 30 35

ggg gac tca gac ccg caa cct gct cct aaa aaa atg aaa aca tct gag 259
Gly Asp Ser Asp Pro Gln Pro Ala Pro Lys Lys Met Lys Thr Ser Glu
40 45 50

tcc tcg acc ata cta gtg gtt cgc tac agg agg aac gtg aaa aga aca 307
Ser Ser Thr Ile Leu Val Val Arg Tyr Arg Arg Asn Val Lys Arg Thr
55 60 65

tct cca gag gaa ctg gtg aat gac cac gcc cga gag aac aga atc aac 355
Ser Pro Glu Glu Leu Val Asn Asp His Ala Arg Glu Asn Arg Ile Asn
70 75 80

```
ccc gac caa atg gag gag gag gaa ttc ata gaa ata acg act gaa aga      403
Pro Asp Gln Met Glu Glu Glu Glu Phe Ile Glu Ile Thr Thr Glu Arg
 85                  90                  95                  100

cct aaa aag tag caagaagcta catccctcaa acttcggcaa tgaaaataaa          455
Pro Lys Lys  *


gtttgagaag ctga                                                      469

<210> 4
<211> 103
<212> PRT
<213> Homo Sapiens

<400> 4
Met Gly Gln Gln Ser Ser Val Arg Arg Leu Lys Arg Ser Val Pro Cys
 1               5                  10                  15
Glu Ser Asn Glu Ala Asn Glu Ala Asn Glu Ala Asn Lys Thr Met Pro
            20                  25                  30
Glu Thr Pro Thr Gly Asp Ser Asp Pro Gln Pro Ala Pro Lys Lys Met
        35                  40                  45
Lys Thr Ser Glu Ser Ser Thr Ile Leu Val Val Arg Tyr Arg Arg Asn
    50                  55                  60
Val Lys Arg Thr Ser Pro Glu Glu Leu Val Asn Asp His Ala Arg Glu
65                  70                  75                  80
Asn Arg Ile Asn Pro Asp Gln Met Glu Glu Glu Glu Phe Ile Glu Ile
                85                  90                  95
Thr Thr Glu Arg Pro Lys Lys
            100

<210> 5
<211> 205
<212> DNA
<213> Homo Sapiens

<220>
<221> misc_feature
<222> (1)...(205)
<223> n = A,T,C or G

<400> 5
cgaccatact agtggttcgc tacaggagga ncgtgaaaag aacatctcca gaggaactgg     60
tgaatgacca cgcccgagag aacagaatca accccgacca aatggaggag gaggaattca    120
tagaaatnac gactnaaagn ccttaaaagt agcaagaagc tacatccctc aaacttcggc    180
aatgaaaata aagtttgaga agctg                                          205

<210> 6
<211> 349
<212> DNA
<213> Homo Sapiens

<400> 6
tggacaaaca atccagtgcc ggcggggtga agaggagcgt cccctgtgaa tccaacgagg     60
tgaatgagac gatgccggag accccaactg gggactcaga cccgcaacct gctcctaaaa    120
aaatgaaaac atctgagtcc tcgaccatac tagtggttcg ctacaggagg aacgtgaaaa    180
gaacatctcc agaggaactg ctgaatgacc acgcccgaga gaacagaatc aaccccctcc    240
```

aaatggagga ggaggaattc atggaaataa tggttgaaat acctgcaaag tagcaagaag 300
ctacatctct caaccttggg caatgaaaat aaagtttgag aagctgaaa 349

<210> 7
<211> 194
<212> DNA
<213> Homo Sapiens

<400> 7
aaaagctcca aaaacctact gtagacatcg aagaaccaat atatacaatg ggccaacaat 60
ccagtgtccg cagctgaaga ggagcgtccc ctgtgaatcc aacgaggcca acgaggccaa 120
tgaggccaac aagacgatgc cggagacccc aactggggac tcagacccgc aacctgctcc 180
taaaaaaaat gaaa 194

<210> 8
<211> 389
<212> DNA
<213> Homo Sapiens

<400> 8
cgaagattca aaacctacaa aagcctgccg cagacattga agaaccaata tatacaatgg 60
acaaacaatc cagtgccggc ggggtgaaga ggagcgtccc ctgtgaatcc aacgaggtga 120
atgagacgat gccggagacc ccaactgggg actcagaccc gcaacctgct cctaaaaaaa 180
tgaaaacatc tgagtcctcg accatactag tggttcgcta caggaggaac gtgaaaagaa 240
catctccaga ggaactgctg aatgaccacg cccgagagaa cagaatcaac ccctccaaa 300
ttgaggagga ggaattcatg gaaataatgg ttgaaatacc tgcaaagtag caagaagcta 360
catctctcaa ccttgggcaa tgaaaataa 389

<210> 9
<211> 473
<212> DNA
<213> Homo Sapiens

<400> 9
gtcaccagga gggtatgcat agggagggca agagctctgg gccactgcga agattcaaaa 60
gctccaaaaa cctactgtag acatcgaaga accaatatat acaatccgcc aacaatccag 120
tgtccgcagg ctgaagagga gcgtcccctg tgaatccaac gaggccaacg aggccaatga 180
ggccaacaag acggtagatg ccggagaccc caactgggga ctcagacccg caacctgctc 240
ctaaaaaaat gaaaacatct gagtcctcga ccatactagt ggttcgctac aggaggaacg 300
tgaaaagaac atctccagag gaactggtga atgaccacgc ccgagagaac agaatcaacc 360
ccgaccaaat ggaggaggag gaattcatag aaataacgac tgaaagacct aaaaagtagc 420
aagaagctac atccctcaaa cttcggcaat gaaaataaag tttgagaagc tga 473

<210> 10
<211> 455
<212> DNA
<213> Homo Sapiens

<400> 10
gtcaccagga gggtatacat agggagggca agagctccgg gccactgaga agattcaaaa 60
cctacaaaag cctgccgctg acattgaaga accaatatat acaatggaca aacaatccag 120
tgccggcggg gtgaagagga gcgtcccctg tgattccaac gaggccaacg agatggtaga 180
tgccggagac ctcgagtggg tactcagacc cgcaacctgc tccgaaaaaa ctaaaaacat 240
ctgagtcctc gaccatacta gtggttcgct acaggaggaa ctttaaaaga acatctccag 300
aggaactggt gaatgaccac gcccgaaaga acagaatcaa ccccctccaa atggaggagg 360
aggaattcat ggaaataatg gttgaaatac ctgcaaagta gcaagaagct acatctctca 420
accttgggca atgacaataa agtttgagaa gctga 455

<210> 11
<211> 45
<212> DNA
<213> Homo Sapiens

<400> 11
taatacgact cactataggg agaaagagga gcgtcccctg tgatt 45

<210> 12
<211> 45
<212> DNA
<213> Homo Sapiens

<400> 12
gatttaggtg acactatagm tacttccatg aattcctcct cctcc 45

<210> 13
<211> 30
<212> DNA
<213> Homo Sapiens

<400> 13
catgccatgg aaagcctgcc actgacattg 30

<210> 14
<211> 30
<212> DNA
<213> Homo Sapiens

<400> 14
catgccatgg acaaacaatc cagtgccggc 30

<210> 15
<211> 7
<212> PRT
<213> Homo Sapiens

<400> 15
Pro Ala Pro Lys Lys Met Lys
1 5

<210> 16
<211> 7
<212> PRT
<213> Homo Sapiens

<400> 16
Pro Lys Lys Met Lys Thr Ser
1 5

<210> 17
<211> 9
<212> PRT
<213> Homo Sapiens

<400> 17
Pro Ala Pro Lys Lys Met Lys Thr Ser
1 5

<210> 18
<211> 21
<212> DNA
<213> Homo Sapiens

<220>
<221> CDS
<222> (1)...(21)

<400> 18
cct gct cct aaa aaa atg aaa 21
Pro Ala Pro Lys Lys Met Lys
1 5

<210> 19
<211> 21
<212> DNA
<213> Homo Sapiens

<220>
<221> CDS
<222> (1)...(21)

<400> 19
cct aaa aaa atg aaa aca tct 21
Pro Lys Lys Met Lys Thr Ser
1 5

<210> 20
<211> 27
<212> DNA
<213> Homo Sapiens

<220>
<221> CDS
<222> (1)...(27)

<400> 20
cct gct cct aaa aaa atg aaa aca tct 27
Pro Ala Pro Lys Lys Met Lys Thr Ser
1 5

<210> 21
<211> 2000
<212> DNA
<213> Homo Sapiens

<400> 21
gcactatgag ccactatgac aggccaaaaa ttttttcttc aaggtctaaa aaagtgtttt 60
cttccatctg tgtaactttc tgtatgtgct tttaaagtcc ttgtgccatt gagttatagg 120
gctttgactt ctgggtctaa aaaggacacc aagttctgct atatcttaaa cactgacagc 180
aattacagcc tcattttcag gcccagtaga agatgctaat caaagtaaac tgcgttcatg 240
agatgcaggg ccagaaatta aaactattca actcctcaag gcccaaggac tatcgtggaa 300
gaggtgggca tgtgagattg taagggctga tattaagaga aaagtagctc agtttctcta 360
ggaattaacc attaatatca aaggcacact aaagcaaaac cagtatctag gttgctgtgt 420

cagtttaaca aggctttctt ggagcattaa ctcactcctt catgcaaaat gatgaaggtt 480
acaaggttta tagaaattat attttatact caagatgatt aaacttttat tatacaattt 540
caaaaaaaca aacttaattt gccgcatcct gtttttaatt agggcttatt gtttgggata 600
ttaagcctcc tctctcaaag aataaagatt ttcacctttt ttttttttg aaacatttga 660
gttactgctt tggttaaatg aatgacttat tttacaatga cccatgaccc tatttcgtga 720
tatcaagcat tttaaacttt ttatctttga cgaactttcc aaagtcaaat tctaacttca 780
ttcctcatta atttttgat atgagtcctc tgaagtccaa aagagacata ttggtcttat 840
ttggtataaa aatcatacaa gaagcatcat caaatataaa atggtgtttg actttctttg 900
ggctgtattt atagaaatgt cattgatatg tgctccaaaa ttatgggaaa ctcttatact 960
tctgaaataa cttgtgtatg ttattaatca taacttttat gttaaattct tgtatgctac 1020
agaagtaacc aaaatttact tgtcaatggt ggttttaata atggccgtcc tggctgggtg 1080
tggtggctca cgcctgtaat cccagcactt tggaggccaa ggcaggtgga tcacgaggtc 1140
aggagttcaa gaccagcttg gccaagatgg tgagaccccg tctctcctaa aaatacaaaa 1200
attagccggg cgtggtggca caggcctgta atcccagcta ctcaggaggc tgagccagag 1260
aatcacttaa acccaggagg cagaggttgc agtgagctga gatcatgcca ctgcactcca 1320
gcctgggcaa caacagtgaa actccatttc aaaaagagaa aaaaatatac acccctgatt 1380
tgaaacaaga ctaggataca cggtaaagaa tgtttattca ctaaagtttt gacagactga 1440
aggccattat ttcagattgg agataggaaa cttaaaagaa ctacatggac attgggtaaa 1500
agaacatggg ttgaaatttg cctggcagta gcccaaaagt aaatggtctt cgaggtgcat 1560
gtgaaggaac tgtaggaggg aaatgggata attcacatct ccgccaataa ataaatgtag 1620
ccacactagc gtgggggatt tgaggtgaga catcgaaaat actaagtcat ggtgagggct 1680
ggaagacaaa agaatgaatc acttcatagg aatggctgtg gggaagggct tgaaggaatc 1740
atcctctgac ttccatgtga accgtgaaca ttaaacatgg agaaatgagg agcggcggca 1800
gatcggtttg ggatgcatct tcaggggatg ctgaaacaac aacagcattt ggtttgctct 1860
acacccctgt cacccctcgc ccgcaagccc agggagtggt cagcagtggg gctttgtgac 1920
gtcgaagcca ccctagggct gccactggct gggacactgc ctgtatgatc aaacaaagct 1980
caagggtgtg gctttgcctt 2000

<210> 22
<211> 2000
<212> DNA
<213> Homo Sapiens

<400> 22
attattataa acttctggat tgcttagaaa aaaaacataa aaaatttttt acttaaggtc 60
attacatctg tgtaactttc tgtatgtgct tttaaagtcc ttgtgccatt gagttatagg 120
gctttgactt ctgggtctaa aaaggacacc aagttctgct atatcttaaa cactgacagc 180
aattacagcc tcatcttcag gcccagtaga agatgctaat caaagtaaac tgcgttcatg 240
agatgcaggg ccagaaatta aaactattca actcctcaag gcccaaggac tatcgtggaa 300
gaggtgggca tgtgagattg taagggctga tattaagaga aaagtagctc agtttctcta 360
ggaattaacc attaatatca aaggcacact aaagcaaaac cagtatctag gttgctgtgt 420
cagtttaaca aggctttctt ggagcattaa ctcactcctt catgcaaaat gataaaggtt 480
acaaggttta tagaaattat attttatact caagatgatt aaacttttat tataaaattt 540
caaaaaaaca aatttcattt gcccatcctg tttttaatta gggcttattg tttgggatat 600
taagcctcct ctctcaaaga ataaagattt tcacctttt ttttttttct tttttgaaat 660
ctttgagtta ctgctttggt taaatgaatg acttatttta caatgaccca tgaccctatt 720
tcgtgatatc aagcatttta aacttttat ctttgacgaa ctttccaaag tcaaattcta 780
acttgattcc tcattaattt tttgatatga gtcctctgaa gtccaaaaga gacatattgg 840
tcttatttgg tataaaaatc atacaagaag catcgtcaaa tataaaatgc tgtttgactt 900
tctttgggct gtatttatag aaatgttatt gatatgtgct ccaaaattat gggaaactca 960
taattctgaa ataacttctg tatgttatta attataaact ttatgttaaa ttgttgtatg 1020
ctacagaagt aaccaaaatt tacttgtcaa tggtggtttt aataatggcc gtcctggctg 1080
ggtgtggtgg ctcacgcctg taatcccagc actttggagg ccaaggcggg tggatcacga 1140
ggtcaggagt tcaagaccag cttggccaag atggtgagac cccgtctctc ctaaaaatac 1200
aaaaattagc cgggcgtggt ggcacaggcc tgtaatccca gctactcagg aagctgagcc 1260
agagaatcac ttaaacccag gaggcagagg ttgcagtgag ctgagatcat gccattgcac 1320
tccagcctgg gcaacaacag tgaaactcca tttcaaaaag agaaaaaaat atacacccct 1380
gatttgaaac aagactagga tacacggtaa agagtgttta ttcactaaag ttttgacaga 1440

ctgaaggcca ttatttcaga ttggagatag gaaacttgaa ctacacggac attgggtaaa 1500
agaacatggg ttgaaatttg cctggcagta gcccaaaagt aaatggtctt cgaggtgcat 1560
gtgaaggaac tgtaggaggg aaatgggata attcacatct ccgccaataa ataaatgtag 1620
ccacactagc ttgggggatt tgaggtgaga catcgaaaat actaagtcat ggtgagggct 1680
ggaagacaaa agaatgaatc acttaatagg aatggctgtg gggaagggct tgaaggaatc 1740
atcctctgac ttccatgtga accgtgaaca ttaaacatgg agaaatgagg agcggcggca 1800
gatcggtttg ggatgcatct tcaggggatg ctgaaacaac agcagcattt ggtttgctct 1860
acacccctgc caccctcgc ccgcaagccc agggagtggt cagcagtggg gctttgtgac 1920
gtcgaagcca ccctagggct gccattggct gggacactgc ctgtatgatc aaacaaagct 1980
caagggtgtg gctttgcctt 2000

<210> 23
<211> 33
<212> DNA
<213> Homo Sapiens

<400> 23
atagttagcg gccgcctugc aggtattuca acc 33

<210> 24
<211> 261
<212> DNA
<213> Homo Sapiens

<400> 24
gctccgaaaa aactaaaaac atttgagtcc tcgaccatac tagtggttcg ctacaggagg 60
aactttaaaa gaacatctcc agaggaactg gtgaatgacc ccgcccgaaa gaacagaatc 120
aacccccctcc aaatggagga ggaggaattc atggaaataa tggttgaaat acctgcaaag 180
tagcaagaag gtacatctct caaccttggg caatgccaat aaagtttgag aagctgaaaa 240
aaaaaaaaaa aaaaaaaaaa a 261

<210> 25
<211> 403
<212> DNA
<213> Homo Sapiens

<400> 25
ggcacgagga gcctgccact gacattgaag aaccaatata tacaatggac aaacaatcca 60
gtgccggcgg ggtgaagagg agcgtcccct gtgattccaa cgaggccaac gagatgatgc 120
cggagacccc aactggggac tcagacccgc aacctgctcc taaaaaaatg aaaacatctg 180
agtcctcgac catactagtg gttcgctaca ggaggaactt taaaagaaca tctccagagg 240
aactgctgaa tgaccacgcc cgagagaaca gaatcaaccc cctccaaatg gaggaggagg 300
aattcatgga aataatggtt gaaatacctg caaagtagca agaagctaca tctctcaacc 360
ttgggcaatg aaaataaagt ttgagaagct gaaaaaaaaa aaa 403

<210> 26
<211> 316
<212> DNA
<213> Homo Sapiens

<400> 26
tcctggcctc tccctgtttt cttaacagat gccggagacc tcgagtgggt actcagaccc 60
gcaacctgct ccgaaaaaac taaaaacatc tgagtcctcg accatactag tggttcgcta 120
caggaggaac tttaaaagaa catctccaga ggaactggtg aatgaccacg cccgaaagaa 180
cagaatcaac cccctccaaa tggaggagga ggaattcatg gaaataatgg ttgaaatacc 240
tgcaaagtag caagaagcta catctctcaa ccttgggcaa ccccaataaa gtttgagaag 300
ctgatggctg tgtaaa 316

<210> 27
<211> 337
<212> DNA
<213> Homo Sapiens

<400> 27
cggggtgaag aggagcgtcc cctgtgattc caacgaggcc aacgagatga tgccggagac 60
cccaactggg gactcagacc cgcaacctgc tcctaaaaaa atgaaaacat ctgagtcctc 120
gaccatacta gtggttcgct acaggaggaa ctttaaaaga acatctccag aggaactgct 180
gaatgaccac gcccgagaga acagaatcaa ccccctccaa atggaggagg aggaattcat 240
ggaaataatg gttgaaatac ctgcaaagta gcaagaagct acatctctca accttgggca 300
atgaaaataa agtttgagaa gctgatggct gtgtaaa 337

THIS PAGE BLANK (USPTO)